# REASONS, AGENCY, AND RESPONSIBILITY: A DEFENCE OF SOFT COMPATIBILISM

WONG SOO LAM

(MA PHILOSOPHY – NATIONAL UNIVERSITY OF SINGAPORE)

A THESIS SUBMITTED FOR THE DEGREE OF DOCTORATE OF PHILOSOPHY

DEPARTMENT OF PHILOSOPHY

NATIONAL UNIVERSITY OF SINGAPORE

2011

**ACKNOWLEDGEMENT**

My utmost gratitude goes to my thesis supervisor Assistant Professor Kyle Swan. I had benefitted tremendously from his meticulous guidance and helpful comments on my thesis drafts, as well as his enthusiastic and timely support when I needed it.

My heartfelt thanks go to the members of my PhD Qualifying Examinations panel – Assistant Professor Kyle Swan, Associate Professor Cecilia Lim, and Assistant Professor Loy Hui Chieh – who had offered insightful suggestions to my thesis proposal. I had also profited greatly from attending Assistant Professor Loy's 'Freedom and Moral Responsibility' course.

Attending the 'Aspects of Responsibility' course (class of 2009) at CEU Summer University conducted by Professor Timothy O'Connor, Professor Mark Balaguer, Professor Michael McKenna, Professor Derk Pereboom, Professor Paul Russell, Professor Thomas Pink, and Dr Andras Szigeti, and participated in by a class of wonderful graduate students, had contributed significantly to my thesis development. I am really thankful for it.

I am very grateful to the Department of Philosophy in NUS for giving me this opportunity to realise my dream of pursuing higher studies in my chosen field, to all the Faculty members whom I had the great opportunity to learn from, to all the administrative staff who had rendered me their kind assistance, as well as to the fellow graduate students who made my stay a joyful and fulfilling one.

Last but not least, I offer my deepest appreciation to my ever understanding and supportive wife Catherine Neo, whose warm encouragement just kept me going.

**CONTENT**

**Chapter Four: Moral Responsibility without Ultimacy Condition**



**Chapter Five: Causal Responsibility Grounds Moral Responsibility**

**Chapter Six: Reasons, Agency, and Responsibility – A Defence of Soft Compatibilism**

**SUMMARY**

I defend a thesis of soft compatibilism. First, I argue that libertarian free will is incompatible with causal determinism and indeterminism because the human will is subject to antecedent conditions. If causal determinism is true, then these antecedent conditions are sufficient for the will to be a certain way. If indeterminism is true, then these antecedent conditions are insufficient for the will to be a certain way. Either way, I argue, certain psychological states with reasons as content figure in the determination or influencing of agents' choices and actions. Second, I argue that a will conditioned by antecedent facts or events is compatible with retrospective moral responsibility, provided that retrospective moral responsibility does not require the ultimacy condition. I argue that it does not. Both determinism and indeterminism allow agents to be retrospectively responsible for their choices and actions because the relevant psychological states with reasons as content figure in the determination or influencing of the agent. Therefore, retrospective moral responsibility is compatible with the will being subject to antecedent conditions, whether or not causal determinism or indeterminism is true.

# CHAPTER 1: AN OVERVIEW OF THE FREE WILL DEBATE AND THE PLACE OF SOFT COMPATIBILISM IN IT

## The Thesis of Soft Compatibilism – An Overview

The Thesis of Soft Compatibilism I aim to argue for is composed of two ideas. First, a constrained notion of freewill (a thesis that agents possess conditioned will) is compatible with the thesis of both causal determinism (where all events have sufficient antecedent conditions) and causal indeterminism (where some events have no sufficient antecedent conditions). Both causal determinism and causal indeterminism allow agents having reasons (or agents exemplifying relevant intentional states with reasons as content) to figure in the co-determination or co-influencing of events. If causal determinism is true, then the agents' choices and actions have sufficient antecedent conditions, which include agents having reasons. If causal indeterminism is true, then the agents' choices and actions have insufficient antecedent conditions, which include agents having reasons. The causal gap in the antecedent conditions can be filled by genuine chance, luck, or random factors. Second, this constrained notion of freewill is compatible with a less robust sense of retrospective moral responsibility. If causal determinism is true, then agents cannot be fully responsible for their choices and actions, because their intentional states with reasons as content themselves have sufficient antecedent conditions. But they can still be partially responsible because they chose and acted based on their reasons, even if their reasons themselves have sufficient antecedent conditions. If causal indeterminism is true, then agents cannot be fully responsible for their choices and actions, because their reasons are not sufficient antecedent conditions for their choices and actions. But they can still be partially responsible because they chose and acted based on their reasons, even if their intentional states with reasons as

content are themselves subject to chance, luck, or random factors. To summarise, a conditioned will is compatible with a less robust sense of retrospective moral responsibility, whether or not causal determinism or causal indeterminism is true.

How is soft compatibilism different from other versions of compatibilism? Classical compatibilism (or soft determinism) is the thesis that free will is compatible with causal determinism, and that causal determinism is true. I have two objections to this view. First, I agree with the incompatibilists' definition of free will, which is the ability to choose and act otherwise given exactly the same antecedent conditions and laws. Defined this way, it is not compatible with causal determinism. If the compatibilists want to say that we can choose and act otherwise only if either the antecedent conditions or laws had been different, it is more accurate to describe it as conditioned will rather than free will. Second, I agree that causal determinism is a plausible thesis for many reasons but I believe that it is yet to be proven true. Contemporary compatibilism is the thesis that causal determinism is compatible with moral responsibility. Most contemporary compatibilists concede that causal determinism is incompatible with free will (the incompatibilists' definition). I have two related objections to this view. First, there are two notions of responsibility – retrospective (backward-looking) and prospective (forward-looking). Some compatibilists hold that causal determinism is compatible with prospective moral responsibility. All incompatibilists can agree with this. I think the crucial question is whether causal determinism is compatible with retrospective responsibility. Some compatibilists say it is not. Second, some contemporary compatibilists who say that causal determinism is compatible with retrospective responsibility either hold that the truth or falsity of causal determinism is irrelevant to moral responsibility, or do not clearly specify the notion of retrospective responsibility. I disagree that the truth or falsity of causal determinism is irrelevant to moral responsibility, for I believe that any causal thesis

(whether deterministic or indeterministic) has significant implications for retrospective moral responsibility. As retrospective responsibility is not clearly specified by the compatibilists, incompatibilists can insist that retrospective moral responsibility requires the ultimacy condition – which is that we are fully responsible for our choices and actions because they are entirely up to us and also because they have their source entirely in us. And the ultimacy condition is hence incompatible with causal determinism, because neither our choices and actions are entirely up to us nor do they have their source entirely in us if causal determinism is true. I agree with the incompatibilists here, but I would add that neither our choices and actions are entirely up to us nor do they have their source entirely in us, even if causal indeterminism is true. In reply, the compatibilists have to specify retrospective moral responsibility as one without the ultimacy condition. That is, they have to say that causal determinism is compatible with retrospective moral responsibility without the ultimacy condition. And this is the position of soft compatibilism. It is soft because both the notions of will and moral responsibility are 'softened' to make them compatible with one another and with either causal determinism or causal indeterminism.

In what follows, I write an overview of the free will debate in general to define my own position against the existing positions in this chapter. And then I state and defend the premises of my position against the premises of some existing positions in more specific areas of the debate in later chapters. The first premise is that the agents' reasons are explanations of action and that reasons explanation is a form of causal explanation. And I attempt to defend a version of the causalist account (where reasons do figure as causes) against the non-causalist (where reasons do not figure as causes) account of reasons explanation in Chapter Two. The second premise is events involving agents are causes of action and that agent causation is a form of event causation. And I attempt to defend a

version of the event causalist account (where agents acting for reasons can be conceived as events) against the agent causalist account (where agents acting for reasons cannot be conceived as events) in Chapter Three. The third premise is that retrospective moral responsibility does not require the ultimacy condition and that the ultimacy condition is both naturalistically implausible and normatively unnecessary. And I attempt to defend retrospective moral responsibility without ultimacy condition from the hard incompatibilists and attack retrospective moral responsibility with ultimacy condition of the libertarians in Chapter Four. The fourth premise is that causal responsibility grounds retrospective moral responsibility, where moral responsibility can be justifiably attributed to the agents' actions and their consequences only when the agents' exemplification of the relevant intentional states with the relevant reasons as content are causally responsible for them. And I attempt to defend the premise against various objections. A concluding essay summarising my position as well as responses to other positions will be the content of Chapter Six. In a nutshell, the thesis of soft compatibilism attempts to establish that an agent exemplifying intentional states with reasons as their content is causally responsible for his choice and action, and hence he is morally responsible for them.

**Determinism, Free Will, and Moral Responsibility – An Overview**

The main argument for hard determinism against freewill can be formulated as follows:

P1. All our choices and actions have sufficient causal conditions. (Determinism Thesis)

P2. If all our choices and actions have sufficient causal conditions, then we do not have free will. (Incompatibility Thesis)

C3. Therefore, we do not have free will. (Denial of the Free Will Thesis)

Assuming this argument is valid, there are three broad theses towards this argument. Hard determinism accepts both premises (1), (2) and conclusion (3). Soft determinists accepts premise (1) and rejects conclusion (3), and hence must reject premise (2). Libertarians accepts premise (2) and rejects conclusion (3), and hence must reject premise (1). Hard determinists agree with soft determinists about premise (1) but disagree about premise (2) and conclusion (3). Hard determinists agree with libertarians about premise (2) but disagree about premise (1) and conclusion (3). Soft determinists agree with libertarians about conclusion (3) but disagree about premises about (1) and (2). There are two more restricted positions in this argument. Hard incompatibilists are uncommitted to premise (1), accepts premise (2) and conclusion (3). That is, conclusion (3) is true no matter whether premise (1) is true or false.

The main argument above for determinism can be extended against moral responsibility as follows:

P3. We do not have free will. (Denial of the Free Will Thesis)

P4. If we do not have free will, then we are not morally responsible for all our choices and actions. (Moral Responsibility Implies Free Will Thesis)

---

C5. Therefore, we are not morally responsible for all our choices and actions. (Denial of the Moral Responsibility Thesis)

Hard incompatibilists accept premise (4) and conclusion (5). That is, conclusion (5) is true no matter whether premise (1) is true or false. Hard compatibilists agrees with hard incompatibilists about premises (1) and (2), as well as conclusion (3), but rejects premise (4)

and conclusion (5). My position is a version of soft compatibilism which agrees with the hard determinists and hard incompatibilists on premises (1), (2), (4) and conclusions (3) and (5). Very briefly, my argument for this is as follows. First, the premise that we do not have free will does not rule out the possibility that we have conditioned will. Second, the premise that we have conditioned will does not rule out the possibility that we are at least morally responsible for some of our choices and actions. Hence, we are still responsible (but not ultimately) for all our choices and actions even if we do not have free will.

**Premise (1): Assessing the Determinism Thesis**

Premise (1) is known as determinism thesis. It states that all our choices and actions are causally determined. An argument for conclusion (1) can be expressed as follows:

P6. All events have sufficient causal conditions.

P7. All our choices and actions are a kind of events.

---

C1. Therefore, all our choices and actions have sufficient causal conditions.

To begin with, premise (6) means that 'every event has a cause', 'all events have causes' or that 'no events are uncaused'. The causes here refer to at least sufficient causes – whenever the cause is present the effect follows, or an event cannot be the sufficient cause of an effect if it sometimes fails to produce the effect. The presence of sufficient causes necessitates their effects. However, it is not mandatory for determinism to maintain that all effects have necessary causes, for it is possible for effects to have different sufficient causes. Determinism is a conditional necessity. That is, a determined event does not have to occur no matter what happens (absolute necessity), but it must occur when the sufficient causal

conditions are present (conditional necessity). Moreover, determinism need not commit itself to what kinds of sufficient conditions are causally relevant. Premise (7) reduces choices and actions into a kind of events. Choices are states of the mind and actions are states of the body, and both are properties exemplified by agents. It is generally accepted that mental states and bodily states exemplified by agents are events. So, premise (1) – the determinism thesis – as derived from premises (6) and (7), says that all our choices and actions are a kind of events, and like other kinds of events, have sufficient causal conditions (including our deliberations over desires and beliefs). Alternative ways to express premise (1) are: past conditions (including our deliberations over desires and beliefs) and the laws of nature determines a unique outcome (including our choices and actions), or given past conditions (including our deliberations over desires and beliefs) and the laws of nature, the outcome (including our choices and actions) cannot be different from what they are (Van Inwagen 1983, pp.16). I believe that this implication of the determinism thesis is crucial in clarifying the debate between the compatibilists and the incompatibilists.

After saying what the determinism thesis entails, it is also useful to say something about what it does not entail. Compatibilists argues that determinism does not imply constraint, control (or manipulation), fatalism, necessary condition, mechanism, predictability, microphysical determinism, irrelevance of proximate cause, or that similar (not identical) past conditions and laws of nature determine a unique outcome. First, free will is not the opposite of determinism. Rather, it is the opposite of coercion, compulsion, and constraint (Kane 2005, pp.18). Coercion, compulsion, and constraint prevent us from choosing and doing what we want and thus act against what we will. Determinism does entail that all events have sufficient causal conditions for their occurrence but it does not necessarily act against our wills and prevents us from choosing and doing what we want. It is a mistake to think that

deterministic causes coerce, compel, or constrain us. Although coercions, compulsions, and constraints can be conceived as the kind of causes that hinder and impede us from choosing and doing what we want, not all causes are of this kind. There is a kind of causes, such as physical strength or volitional power that enables us to choose and do what we want. So, whether our choices and actions are free or not depends on what kinds of causes they are. While some enhance our freedom of choice and action, others hinder or impede it. Second, determinism is not control (or manipulation) by other agents (Kane 2005, pp.19). Since nature (even if it is deterministic) is not an agent, it does not control or manipulate us. So, while free will is the opposite of control, it is not necessarily the opposite of determinism. Third, determinism is not fatalism (Kane 2005, pp.19). Fatalism implies absolute necessity – the view that whatever is going to happen must happen no matter what we choose or do. By contrast, determinism implies conditional necessity – the view that what we choose or do would make a difference in whatever is going to happen, even if what we choose or do themselves have sufficient causes.

Fourth, determinism is about events having sufficient conditions and not about events having necessary conditions (Taylor and Dennett 2002, pp.271). This may seem to contradict the earlier statement that determinism is conditional necessity but it does not on closer scrutiny. Conditional necessity means that if a set of conditions, say X, Y, Z, that are jointly sufficient for event A are present, then event A necessary follows. That is, a set of jointly sufficient conditions X, Y, Z necessitate event A. But this does not preclude event A from having other sets of jointly sufficient conditions, say, X1, Y1, Z1. Hence, neither the set of conditions X, Y, Z, nor the set of conditions X1, Y1, Z1, is necessary set of conditions of event A. The 'necessity' in conditional necessity refers to the necessitation relation between a set of jointly sufficient conditions and an event. It implies nothing about whether any set, or any member

of a set, is a necessary condition. And it implies nothing about what kind of sufficient conditions or jointly sufficient conditions are there. Fifth, determinism is not mechanism (Kane 2005, pp.20). To say that our choices and actions are determined is not to say that we are all like machines (robots or computers) running mechanically on a fixed set of rules, or like lower animals (amoebae and insects) responding automatically with a narrow set of responses to the stimuli from our environment. Even if determinism is true, there is still vast difference between humans and machines on one hand, as well as humans and lower animals on the other. Unlike machines, humans are highly complex creatures running on a complex set of rules that are practically unpredictable. And unlike lower animals, humans respond to the stimuli of our environment with a wide array of responses after deliberation over reasons, reflection on past events, and anticipation of possible future events. This remains true despite our deliberations, reflections, and anticipations having sufficient causal conditions which include our character, personality, and motivational structure. Sixth, determinism need not imply microphysical determinism. That is, it need not commit to the thesis that all causal conditions can only be given at the level of the basic science – physics; it is consistent with the thesis that causal conditions can also be given at the level of the special sciences – chemistry, biology, psychology etc.

Seventh, determinism does not imply the causal irrelevance of proximate causes (Van Rensselar 1958, pp.231). Suppose A, B, C, D, E are a chain of events, whereby the preceding event in the series is a sufficient causal condition for the succeeding one. If we want to find out the sufficient causal conditions for E, A would the remotest cause and D would be the most proximate cause. Remoteness and proximity are usually understood in (but not necessarily limited to) the spatiotemporal sense. The fact that A is the sufficient causal condition for B, and similarly B for C, C for D, and D for E, does not imply that A alone is a

sufficient causal condition for E. Rather, A, B, C and D in conjunction are the sufficient causal conditions for E. A, B, and C do not render D causally irrelevant to E. Hence, D remains causally relevant to E even if it has prior sufficient causal conditions A, B, and C. Eighth, determinism does not imply that similar past conditions (including our deliberations over desires and beliefs) and the laws of nature do determine a unique outcome (including our choices and actions), or that given similar past conditions (including our deliberations over desires and beliefs) and the laws of nature, the outcome (including our choices and actions) cannot be different from what they are. However, it does imply that identical past conditions and the laws of nature do determine a unique outcome (including our choices and actions), or that given identical past conditions and the laws of nature (including our deliberations over desires and beliefs), the outcome (including our choices and actions) cannot be different from what they are (Walter 2009, pp.42). Incompatibilists may agree with the compatibilists about what determinism does not entail but disagree about what free will amounts to. For the incompatibilists, the free will thesis implies the following: that identical past conditions and the laws of nature do not determine a unique outcome (including our choices and actions); or that given identical past conditions and the laws of nature, the outcome (including our choices and actions) can be different from what they are. Again, I believe that this implication of the free will thesis is crucial in clarifying the debate between the compatibilists and the incompatibilists.

Assuming the above argument for premise (1) is valid, libertarians who want to deny premise (1) have to show that either premises (6) or (7), or both, are false. Denial of premise (6) implies affirmation of some form of indeterminism. Libertarians who take this option are usually known as simple indeterminists or causal indeterminists. Most of them accept premise (7) but insists that some events (like choices and actions) are uncaused or

indeterministic. Moreover, there is a general consensus among philosophers that premise (6) can neither be proven nor disproven, and thus the causal indeterminism thesis remains a possibility. Denial of premise (7) implies that choices and action are not events. Libertarians who take this option are usually known as the agent causalists. They are able to accept premise (6) but insist that unlike events, agents (who choose and act) are not causally necessitated by past conditions and the laws of nature. For while events are mere occurrences and happenings, agents are the initiators, origins, or sources of their choices and actions. Both options, however, have their own problems. I will look into these problems in greater detail later. My soft compatibilist position remains uncommitted towards premise (6) but affirms premise (7). That is, it is not committed to the truth of causal determinism with respect to events in general but affirms that choices and actions can be conceived as a kind of events. So, it follows that my position is not committed to premise (1).

**Premise (2): Assessing the Incompatibility Thesis**

Premise (2) is known as the incompatibility thesis. It simply states that causal determinism and free will are incompatible or mutually exclusive. That is, if one is true, the other must be false. They cannot be both true. Incompatibilists offer various arguments for premise (2), which usually has the following form (Thornton 1989, pp.45-7):

a. If the free will thesis is true, then X is the case.

b. If the determinism thesis is true, then X is not the case

---

c. Therefore, if determinism is true, then we do not have free will.

This argument is valid as long as X means the same thing in both premises (a) and (b). Two comprehensive ways of specifying X has already been discussed above. Both are expressions of causal indeterminism. The first way is 'X = past conditions and the laws of nature do not determine a unique outcome (including our choices and actions)'. The second way is 'X = given past conditions and the laws of nature, the outcome (including our choices and actions) can still be different from what they are'. A more common way to specify X focuses on our choices and action, which is 'X = we can act otherwise than we do act'. With this in mind, this argument in favour of premise (2) can be formulated as follows:

P8. If we have free will, then we can act otherwise than we do act.

P9. If all our choices and actions have sufficient causal conditions, then it is not the case that we can act otherwise than we do act.

---

P2. If all our choices and actions have sufficient causal conditions, then we do not have free will.

To reject the conclusion of this argument, compatilibilists have to either show that the argument is not valid or that at least one of the premises is false. As mentioned, arguments of this form are valid only when 'X = we can act otherwise than we do act' means the same thing in both premises (8) and (9). So, the compatibilists have to show either that X means a different thing in premises (8) and (9) and the argument is invalid because of the fallacy of equivocation, or that X has the same meaning in premises (8) and (9) and at least one of the premises is false. The compatibilists reads 'we can act otherwise than we do act' as 'if we choose to act otherwise, then we will' in premise (8). This is known as the hypothetical reading of 'can act otherwise' as opposed to the categorical reading ('it is in our power to act otherwise than we do') adopted by the incompatibilists. The hypothetical reading of X

preserves the truth of premise (8) but renders premise (9) false. This is because it is compatible with determinism to say that if we had chosen differently, then we would have acted differently from the way we actually did. To preserve the truth of premise (9), X would have to mean something different. If this the case, then this argument would become invalid because of the fallacy of equivocation. For it is not possible for X to mean the same thing (under the hypothetical reading) in both premises (8) and (9). In reply, the incompatibilists insist on the categorical reading of 'can act otherwise'. If our choices and actions are causally necessitated in accordance with deterministic laws, then it is not in our power to choose or act otherwise than we do because it is impossible to avoid deciding and doing what follows necessarily from deterministic laws. The categorical reading of X means the same thing in premises (8) and (9), and hence the argument does not commit the fallacy of equivocation. The compatibilists respond by saying that being able to choose and act otherwise means that if one prefers an alternative course of action, then one would choose and act on that alternative. That is, one would choose and act differently only if some past condition (preference for an alternative course of action) had been different. The incompatibilists retort that if our choosing or acting otherwise implies a difference in some antecedent conditions, then it is not in our power to choose or act differently given the past conditions exactly as they are. For the past conditions will always make it impossible for us to choose or act differently. The compatibilists rejoin by saying that determinism does not make different outcomes impossible, it just implies that different outcomes always involve different past conditions.

The standard incompatibilists' argument for premise (9) can be summarised as follows: according to determinism, past conditions and the laws of nature imply that we will choose or act in a certain way. If we are able to choose or act otherwise, then we must either be able to

change the past conditions or the laws of nature. Since we are not able to change the past conditions or the laws of nature, we are not able to choose or act otherwise. And the standard compatibilists' argument against premise (9) can be summarised as follows: according to determinism, past conditions and the laws of nature imply that we will choose or act in a certain way. If we choose or act otherwise, then either the past conditions or the laws of nature must have been different. And if this is the case, then our choosing or acting otherwise does not imply changing the past conditions or the laws of nature. It seems to me that the main disagreement between the incompatibilists and compatibilists is over what free will really means. The freedom of the will is categorical (alternative choices and actions are possible given exactly the same past conditions and laws of nature) for the incompatibilists and hypothetical (alternative choices and actions are possible only if the past conditions or the laws of nature had been different) for the compatibilists. There is a general consensus among philosophers that the incompatibilists and compatibilists have reached an impasse or stalemate here. To break the impasse or stalemate, I suggest conceding premise (2) to the incompatilists for purely semantical reasons. In agreement with the incompatilists, a will that is free cannot be itself caused or conditioned by the past or the laws of nature. That is, free will refers to unconditioned will. The freedom to choose or act otherwise seems to imply not only freedom from constraints, but freedom from conditions as well. To avoid confusion, my soft compatibilist position asserts that it is conditioned will, and not free will, which is compatible with the determinism thesis. We should reserve the term free will to mean the unconditioned will. It also argues that (a) affirmation of determinism makes free or unconditioned will impossible and that (b) denial of determinism (or affirmation of indeterminism) does not necessarily imply free or unconditioned will.

**Conclusion (3): Assessing the Denial of Free Will Thesis**

Before assessing the main arguments for the free will thesis, it is useful to discuss and put aside the weaker arguments first. The first one appeals to the phenomenology of free will (Thornton 1989, pp.105). It argues from the mere premise that we are conscious of being free to the conclusion that we have free will. But this argument is mistaken. Even if we feel that we are free in our choices and actions, our choices and actions may be causally necessitated by conditions unknown to us. Hence, we cannot derive the conclusion that we have free will from the mere premise that 'we are conscious of our freedom' or 'we feel that we are free', for it would be a phenomenological fallacy to argue from a phenomenological (what appears to be the case) premise to an ontological (what is the case) conclusion. The second one is related to the first. It argues from the premises that (a) choices always involves alternatives and that (b) we can choose among these alternatives to the conclusion that we have free will (Thornton 1989, pp.109). The objection to the first argument applies to the second. Similarly, there may be unknown antecedent sufficient conditions for us to choose one alternative over another. The third argument asserts that all our choices and actions involve some non-physical element like the mind, soul, or spirit (Kane 2005, pp.40-4). A well-known difficulty with this view is that human beings, who are capable of making choices and taking actions, are physical beings subject to the physical laws of nature. Some libertarians may want to appeal to Descartes' dualism and insist that the mental is separate and distinct from the physical and hence it is not subject to the physical laws of nature. This claim is highly contentious if not rejected outright nowadays. For libertarians who assert it have to account for how the mental causally interact with the physical, given that the physical can be completely explained by the physical laws of nature. Another version of this claim is Kant's distinction between autonomy and heteronomy. Where the latter belongs to the world

of senses (empirical / phenomena) and is governed by laws of nature, the former belongs to the world of understanding (rational / noumena) and is governed by the laws of reason. More pointedly, the latter is not subjected to the laws of the former. The Kantian claim faces a difficulty similar to the Cartesian claim. Analogously, libertarians who assert it have to account for how the autonomous causally interact with the heteronomous, given that the heteronomous can be completely explained by heteronomous laws of nature. The fourth argues from the premise that our choices and actions are not predictable to the conclusion that we have free will and that the determinism is false. Determinism does imply predictability in principle because if a set of causes sufficient to bring about a certain effect exist, then one can predict the effect by knowing the jointly sufficient set of causes. In other words, if one knows the exact past conditions and all the relevant laws of nature, then one can predict the future outcome. However, determinism is consistent with practical unpredictability. It is possible that we cannot know the exact past conditions and all the relevant laws of nature, and the future outcome remains practically unpredictable even if determinism is true. So, unpredictability implies neither the falsity of determinism nor the truth of free will. And as we shall see later, the falsity of determinism does not imply the truth of free will as well.

Let's move on to the main arguments for the free will thesis. Analogous to the structure of the main argument for hard determinism against free will, the first main argument for freewill can be formulated as follows:

P10. All our choices and actions have no sufficient causal conditions. (Indeterminism Thesis)

P11. If all our choices and actions have no sufficient causal conditions, then we have free will. (Second Incompatibility Thesis)

P12. Therefore, we have free will. (Affirmation of the Free Will Thesis)

But unlike the main argument for hard determinism against free will, the premises of the main libertarian argument for free will are too weak to support the conclusion. In other words, the conjunction of the indeterminism thesis and the incompatibility thesis are insufficient to establish the truth of the free will thesis. An argument for premise (10) can be expressed as follows:

P6. All events have sufficient causal conditions.

P13. None of our choices and actions is an event.

---

P10. Therefore, all our actions and choices have no sufficient causal conditions.

Assuming that this argument is valid, it seems that more support is required to support premise (13) against premise (7) discussed earlier, to favour premise (10) over premise (1). Although premises (10) and (11) are causally insufficient for premise (12), libertarians can still plausibly claim that the premise (10) – the indeterminism thesis – is a necessary but not sufficient causal condition for the free will thesis. But even if this is true, additional premises are still required to establish the free will thesis. So, while premise (2) – if determinism is true, then we do not have free will – could be true, it's converse, premise (11) – if determinism is false, then we have free will – is obviously false. The fact that all our choices and actions have no sufficient causal conditions does not imply that we have free will. For it remains possible our choices and actions that have no sufficient causal conditions are uncaused or random. So, premise (11) should be rejected. Here again, the falsity of determinism does not imply the truth of free will.

**Premise (4): Assessing the Moral Responsibility Implies Free Will Thesis**

Premise (4) is stated above as follows: if we do not have free will, then we are not morally responsible for all our choices and actions. Another way to state premise (4) is that free will is a necessary condition of moral responsibility and that moral responsibility is a sufficient condition of free will. Premise (4) can then be reformulated as premise (15) below: if we are morally responsible for all our choices and actions, then we have free will. With this in mind, the second main argument for free will can be expressed as follows:

P14. We are morally responsible for our choices and actions.

P15. If we are morally responsible for our choices and actions, then we have free will.

---

P12. Therefore, we have free will.

This seems to be the master argument for the free will thesis. It takes moral responsibility for our choices and actions as a premise to imply that we have free will. Its strong intuitive appeal lies in the fact the premise (14) can justify most if not all of our moral and legal practices. An objection would be that we cannot derive the conclusion that 'we have free will' from the mere premise that 'we are morally responsible for our choices and actions', for it would be a deontological fallacy to derive a positive/ontological conclusion (what is the case) from a normative/moral premise (what ought to be the case) (Walter 2009, pp.59). If this is the only argument that libertarians rely on to establish the free will thesis, then its validity would be questionable. But most libertarians do offer independent positive/ontological premises in support of the free will thesis, like P10 and P13 discussed above. It is more plausible to establish positive/ontological premises to derive positive/ontological conclusions first, and find out what normative/moral theses they can support next. So, even if our moral practices are usually deemed too important to be given

up, they cannot be used as premises to derive ontological theories. Having said that, premise (14) need not be denied. It may still be supported by some existing positive/ontological theses. I concede premise (14) to the libertarians with a qualification that we are only morally responsible for some but not all of our choices and actions. I believe that how much of our choices and actions we are morally responsible for remains an open question that deserves further analysis. Premise (15) is contentious. For compatibilists, the fact that we are morally responsible for our choices and actions does not imply that we have free will, simply because they hold that moral responsibility is compatible with determinism. This is so because our desires, believes, intentions, and reasons are part of the sufficient causal conditions leading to our choices and actions, and this is compatible with causal determinism. The incompatibilists and the compatibilists arrive at another impasse or stalemate here. Most libertarian arguments eventually fall back on some version of the moral responsibility argument to justify their metaphysics of free will. As discussed above, the word 'free' for incompatibilists implies not merely 'unconstrained' but also 'unconditioned'. Read this way, premise (15) implies that we can only be morally responsible for our choices and actions only if our will is unconditioned (or free from all antecedent causal conditions). The compatibilists object that if our choices and actions have no antecedent causal conditions at all, then our desires, beliefs, intentions, and reasons (which are perhaps grounded in our character, personality traits and motivational structure) are not at least in part antecedent causal conditions of our choices and actions. And if our desires, beliefs, intentions, and reasons (which are perhaps grounded in our character, personality traits and motivational structure) are not at least in part antecedent causal conditions of our choices and actions, then how can we be morally responsible for them at all. I agree with the compatibilists here. My soft compatibilist view holds that being morally responsible for some of our choices and actions does not necessarily imply (or even require) that we have free will. It does however

imply that we have conditioned will. We are morally responsible for some of our choices and actions because our will is conditioned (whether sufficiently or not) by our desires, beliefs, intentions, and reasons (which are perhaps grounded in our character, personality traits and motivational structure). That is, I accept premise (14) with qualification and reject premise (15), and hence reject premise (12). And as I remain uncommitted to the truth or falsity of premise (1) – determinism thesis, I remain uncommitted to the truth or falsity of premise (10) – indeterminism thesis too.

**Conclusion (5): Assessing the Denial of Moral Responsibility Thesis**

Both the hard determinists and hard incompatibilists arrive at conclusion (5) by accepting conclusion (3) and premise (4). The libertarians reject conclusion (5) by rejecting conclusion (3) and accepting premise (4). And they reject conclusion (3) by rejecting premise (1) and accepting premise (2). The soft determinists also reject conclusion (5) by rejecting conclusion (3) and accepting premise (4). But they reject conclusion (3) by accepting premise (1) and rejecting premise (2). The hard compatibilists agree with the hard incompatibilists on Premises (1), (2) and conclusion (3), but rejects premise (4) and hence rejects conclusion (5). My version of soft compatibilism actually agrees with the hard determinists and hard incompatibilists on premises (1), (2), (4) as well as conclusions (3) and (5). My disagreement with the hard incompatilists and hard determinists is in the treatment of premises (2) and (4). On premise (2), I agree that if all our choices and actions have sufficient causal conditions, then we do not have free will. But I claim that even if all our choices and actions have sufficient causal conditions, we do have conditioned will. On premise (4), I disagree that if we do not have free will, then we are not morally responsible for our choices and actions at all. For I claim that if we do have conditioned will, then we are

still morally responsible for some of our choices and actions. Will and moral responsibility do not have to be either all or nothing concepts. The thesis that ultimate moral responsibility implies free will, or that free will is necessary for ultimate moral responsibility may be true. But this stronger thesis does not rule out the weaker thesis that non-ultimate (adequate) moral responsibility implies conditioned will, or that conditioned will is necessary for non-ultimate (adequate) moral responsibility.

**References:**


Clarke, Randolph (2009) 'Incompatibilist (Nondeterministic) Theories of Free Will', *Stanford Encyclopedia of Philosophy* (SEP) http://plato.stanford.edu/entries/incompatibilism-theories

Kane, Robert (2005) *A Contemporary Introduction to Free Will* (Oxford University Press)

McKenna, Michael (2009) 'Compatibilism', *Stanford Encyclopedia of Philosophy* (SEP) http://plato.stanford.edu/entries/compatibilism

O'Connor, Timothy (2005) 'Free Will', *Stanford Encyclopedia of Philosophy* (SEP) http://plato.stanford.edu/entries/freewill

Taylor, Christopher and Dennett, Daniel (2002) 'Who's Afraid of Determinism? Rethinking Causes and Possibilities' in *The Oxford Handbook of Free Will* ed. Robert Kane (Oxford University Press)

Thornton, Mark (1989) *Do We Have Free Will?* (Bristol Classical Press)

Strawson, Galen (2005) 'Free will' in *The Shorter Routledge Encyclopedia of Philosophy* ed. Edward Craig (Routledge)

Timpe, Kevin (2006) 'Free Will', *Internet Encyclopedia of Philosophy* (IEP) http://www.iep.utm.edu/freewill

Van Inwagen, Peter (1983) *An Essay on Free Will* (Oxford University Press)

Wilson, H. Van Rensselar (1958) 'On Causation' in *Determinism and Freedom* ed. Sidney Hook (New York University Press)

Vihvelin, Kadri (2007) 'Arguments for Incompatibilism', *Stanford Encyclopedia of Philosophy* (SEP) http://plato.stanford.edu/entries/incompatibilism-arguments

Walter, Henrik (2009) *Neurophilosophy of Free Will* (MIT Press)

## CHAPTER 2: REASONS EXPLANATION AS CAUSAL EXPLANATION

### Introduction

In this essay, I seek to critically evaluate the simple indeterminist or non-causalist account of reasons explanations of action and to defend a causalist account within the context of the free will debate. On the simple indeterminist or non-causalist view, I refer mainly to Carl Ginet's paper 'Reasons Explanations of Action: Causalist versus Non-causalist Accounts'. Central to Ginet's non-causalist claim is that reasons explanations of action imply no causal connection between reasons and actions. On his non-causalist account, an action is free if it is undetermined, done for a reason or purpose, and possesses an actish phenomenal property. He remains silent on the causes of action and insists that the non-causalist account of reasons explanation remains valid in the presence of causes of action. Contra Ginet, I attempt to argue that reasons explanations of action imply causal connection between reasons and action. I take action to mean roughly bodily movement and intentional action to mean roughly voluntary bodily movement even if they are not considered strict equivalents. My view is consistent with the difference-making, counterfactual-supporting contributory, and J.L. Mackie's INUS (Insufficient but Non-redundant part of an Unnecessary but Sufficient) condition accounts of causation. And my main claim is that if reasons explain intentional actions, then reasons must make a difference, counterfactually support, contribute towards, or are INUS conditions of voluntary bodily movement. In other words, reasons have to figure amongst the causes of intentional actions in order to explain them.

Donald Davidson proposes deterministic causalism, where action has deterministic causes. Robert Kane proposes indeterministic causalism, where action has indeterministic causes.

My own version of causalism remains neutral between deterministic and indeterministic causalism, and aim to establish conditioned (not free) choice and action instead. It is not committed to the view that our reasons are the sole causes or sufficient causal conditions of our actions. Rather than conceiving reasons and purposes as isolated enduring states, my view conceives agents exemplifying enduring intentional states, with reasons and purposes as its content or object at certain times (or durations of time), as a single event constituting a single cause which forms at least a part of a set of sufficient causal condition of actions. I will also examine and briefly respond to the main objections against causalism, and in doing so, explicate my own version of causalism even further. The objections discussed include (1) action does not necessarily require voluntary bodily exertion, (2) the logical connection between intention and action, (3) deviant causal chains, and (4) teleological explanation as non-causal explanation. Under (4), the views of Scott Sehon, Stewart Goetz, Hugh McCann, and George Wilson are examined and responded to.

**Simple Indeterminism or Non-Causalism**

According to simple indeterminists, there are two ways to explain events – in terms of causes and in terms or reasons and purposes. When explained in terms of causes, events are happenings and occurrences. And when explained in terms of reasons and purposes, events are actions. Some actions are free. And free actions are uncaused events, but this does not imply that free actions occur by chance. They are not random because though free actions are uncaused events, they can be explained by the reasons and purposes of agents. The first problem the simple indeterminists have to address is that if free actions are uncaused events, what makes them actions and not mere happenings and occurrences? Carl Ginet (2002), a prominent simple indeterminist, answers that an action begins with a simple mental act, a

volition or act of will that initiates the action. This differentiates actions from mere happenings or occurrences. The second problem is what makes events initiated by the volition or act of will actions rather than happenings or occurrences? Ginet answers by saying that a volition or act of will has a certain actish phenomenal quality (Ginet 2002) directly experienced by agents as something they are doing rather than as something happening or occurring to them. The third problem is that an event having actish phenomenal quality may only show that it is an action, but it does not show that it is a free action. Ginet agrees that actish phenomenal quality only differentiates actions from mere happenings or occurrences but it never tells us whether actions are free are not. For actions to be free, they must not only possess an actish phenomenal quality, but they must also be undetermined and done for a reason or purpose.

**Reasons Explanation as Non-Causal Explanation of Action**

As mentioned above, the simple indeterminists distinguish between two types of explanations – causal and reasons. Causal explanations apply to happenings and occurrences while reasons explanations apply to actions. They claim that reasons explanations are not causal explanations, or that reasons are not causes. This is known as the non-causalist account of reasons explanations, defended by Ginet (Ginet 2002) among others. An opposing position claims that reasons explanations are causal explanations, or that reasons are causes. This is known as the causalist account of reasons explanations, defended by Donald Davidson (Davidson 1980) among others. Davidson argues for causalism by pointing out that agent having a certain reason to perform a certain action does not entail that he performed it for that reason. In other words, (P1) "S did A and S had reason R for doing A" does not entail (P2) "S did A because S had a reason R for doing A". And he contends that for (P1) to entail (P2),

we need to add (P3) "S's having reason R caused S's doing A". That is, for (P1) to entail (P2), the explanatory connection between the reason and the action that must be constituted by a causal connection (P3) between them. Davidson believes that the best argument for causalism is that it alone promises to give an account of the explanatory connection between reasons and actions (Davidson 1980, pp.9-12).

Non-causalists, like Ginet, disagree that there are no satisfactory alternatives. Ginet believes that for any true reasons explanation, we can formulate a sufficient truth condition for it without entailing that the explanans (reasons) caused the explanadum (actions). He uses reasons explanation of the form (P4) "S A-ed in order to B" as an example. (P4) can be paraphrased as "S's purpose / intention / aim in A-ing were to B". A sufficient truth condition for (P4) can be given as (P5) "Concurrently with her A-ing S intended of that A-ing that by it (and in virtue of its being an A-ing) she would B (or would contribute to her B-ing)". (P5) guarantees the truth of (P4) without giving causal conditions. It does not entail that the accompanying intention caused the action. And it does not entail anything at all about what caused the action (or any physical or mental ingredients) in the action. Yet (P5) is sufficient for the truth of (P4) (Ginet 2002, pp.388).

As discussed by Ginet, causalists like Alfred Mele disagree with this claim. Mele thinks that (P5) is not a sufficient truth condition for (P4), unless a causal condition is added, and so (P5) can be revised as (P6) "Concurrently with her A-ing S intended of that A-ing that by it (and in virtue of its being an A-ing) she would B (or would contribute to her B-ing) and the intention caused her A-ing". Mele argues that if we can imagine the neural realisation of a particular concurrent intention plays no causal role in the production of action, then we have imagined a situation in which the presence of the intention does not contribute to explaining the action. And Ginet criticises Mele's argument for assuming that neural realisations of

concurrent intention play a causal role in the production of action. While conceding that the assumption may be true, it has not been proven so. But even if we are ignorant of the truth of this assumption, we are not ignorant of the truth of the (non-causal) reasons explanations we confidently give (Ginet 2002, pp.389-90).

**Non-Causalist Arguments against Deterministic Causalism**

Ginet believes that one objection that Davidson did not deal with effectively is this: The explanans in reasons explanations often include enduring states of the agent, like beliefs, desires, and intentions. But according to Davidson, a causal explanation of an event requires that there be antecedent event among the totality of relevant causal factors. He requires this because he assumes that causal laws are deterministic, specifying that when a certain collection of causal factors obtain at a certain time, a certain effect would ensue. Not only must the totality of relevant causal factors explain why a certain event occurred, but the timing of those causal factors obtaining must also explain when a certain event occurred. The objection is that no relevant causal factor in reasons explanations seems to play that role. According to Ginet, there is only a certain combination of reason-states that obtained for some interval and that could have led to the explained action at any of the several times during that interval. But for Davidson, events can be closely associated with reason-states (as relevant causal factors). In his case of a driver who signals a turn by raising his arm, at some moment the driver noticed (or thought he noticed) his turn coming up, and that is the moment he signalled. Ginet objects that the moment the driver noticed his turn coming up might not have been the moment he signalled, for he might have signalled at some other moments. And no matter which moment he signalled, the reasons explanation of the driver raising his arm has been the same – he wanted to let others know that he wanted to turn at the next

intersection and believed that raising his arm would accomplish that. It is implausible to suggest that, as a matter of causal law, the driver's notice of his turn coming up (or some neural realisation of this event) would, in sufficiently similar circumstances, always cause his signalling after exactly the same interval as occurred in this case. Davidson acknowledges that there seem to be cases of intentional action where we cannot explain at all why we acted when we did, parallel to explaining the collapse of the bridge from a structural defect. He suggests that it is because in both cases, we are ignorant of the event or sequence of events that caused the intentional action, and not because there is no such event or sequence of events. Ginet replies that even if Davidson is right about the bridge case (a case of causal explanation), it is not at all evident that the reasons case is parallel, that the truth of reasons explanation requires some event or set of events to explain the timing of the action; and Davidson thinks they are parallel because he already assumes that reasons explanations must be deterministic causal explanations (Ginet 2002, pp.394-5).

**Non-Causalist Arguments against Indeterministic Causalism**

Unlike deterministic causalism, its indeterministic counterpart does not face the problems encountered by Davidson. This is because indeterministic causalism assumes that causal laws are indeterministic, specifying that when a certain collection of causal factors obtain at a certain time, there is a chance that a certain effect would ensue. And the totality of relevant causal factors may explain why a certain event is likely to occur, but it may not explain when a certain event occurred. Ginet notes that indeterministic causalists, like the non-causalists, think that the truth of determinism is incompatible with free action and wish to show how their incompatibilism does not entail the absurd conclusion that no free action has a reasons explanation, or that no one can act freely and responsibly and at the same time act for a

reason. He mentions Robert Kane as an indeterministic causalist, who holds that reasons do play a causal role in the agent's choices and actions but they need not determine it, and that indeterministic event causation of a choice or action by the agent's 'self network' (character and motives) is essential if the choice or action is to be something the agent causes, rather than something that just happens to him. Ginet objects that Kane gives no argument for this claim. While both of them reject the assumption that an agent's control over an action requires that the action has an antecedent determining cause, Ginet rejects the assumption that an agent's control over an action requires that the action has an antecedent non-determining cause. He believes that even if causal indeterminism fits best with science's understanding of physical processes (including perhaps the neural realisations of reason-states), it does not follow that causal indeterminism fits best with reasons explanations; and adds that even if we were to discover that our neural processes underlying our deliberated decisions offers no support for us to claim that our decisions are caused (whether deterministically or indeterministically) by our reasons for them, this would not be ground for claiming that we do not after all make our decisions for those reasons (Ginet 2002, pp.395-7).

**Non-Causalist Account of Intention-Action Connection**

If the relation between intention and action is not causal, what else can it be? According to Ginet, there are two relations – internal and external. On the internal relation, the connection follows from the intrinsic properties of the relata. On the external relation, the connection lies in the (intentional) content's direct reference to the action, or the (intentional) content with the action as its direct referent. The connection consists of these internal and external relations between intention and action, and not a causal relation. On the connection between the internal and external relations, since the agent is directly aware of at least the initial

conscious volitional part of the particular action, it seems that she can refer directly to that action, in a demonstrative fashion – 'this action'. Perhaps this requires that something intrinsic to the intention is caused by part of the action. If this is the case, then it is obvious that this will not entail that the intention cause any part of the action because the causation is in the opposite direction.

Ginet elaborates on the difficulty of conceiving the relation between intention and action as a causal connection: if (a) direct reference requires that what is referred to – the action – causes what refers to it – the intention, and if (b) causes must always precede their effects, then (c) there will be a small period at the beginning of the action during which the concurrent directly referring intention will not be in place, a gap during which the agent will have no intention about the action already begun, during which it will not be a case that she intends that action. (c) is implied by (a) and (b). It is a very unattractive consequence. To reject (c), we have to reject either (a) or (b) or both. Both (a) and (b) seem reasonable to Ginet if the connection is assumed to be causal. For (a), it is difficult to accept that direct reference does not involve the direct referent's entering into causing that which does the referring. For (b), it is difficult to accept the possibility of an efficient cause that occurs simultaneously with (if not later than) the effect. Ginet suggests that (c) can be rejected if the connection in (a) is non-causal, for not all direct demonstrative reference require a causal relation. In his example, if one can simultaneously begin uttering an expression of an intention that directly refers to a concurrent action, without any part of the action's causing any part of the utterance, then one can simultaneously begin having an intention that refers to the action without any part of action causing any part of the intention.

Some philosophers suggest that the notion of intention guiding action is a causal one. When an intentional action involves the voluntary exertion of the body, the agent intends of the

voluntary exertion, under some description of it, that it will accomplish the intended action. Ginet agrees that it seems right to say that the intention, in virtue of having this content, guides the course of the voluntary exertion, or rather that the agent is guided by it in making the voluntary exertion so as to conform to its content. If this is right, then there is a causal connection involved between intention and action after all, for guidance is a causal notion. But Ginet points out that sometimes guidance is a causal notion and sometimes it is not. What makes it the case that an antecedent intention guides a voluntary exertion is not its causing the exertion. Rather, it is the agent's concurrently intending of the voluntary exertion that it conforms to the content of the antecedent intention. The only causal notion required is whatever causation is involved in the agent's currently remembering the antecedent intention and its content, and this does not entail that the antecedent intention causes either the voluntary exertion or the concurrent intention about it. The series of concurrent intentions makes the explanatory link between the antecedent intention and the whole voluntary exertion, but these concurrent intentions that provide the explanatory link do not causally produce the voluntary exertions they accompany. The explanatory link is made simply by the content of the concurrent intentions, including their direct references to the antecedent intention and to the concurrent actions (Ginet 2002, pp.390-1).

**Reasons Explanation as Causal Explanation of Action**

To evaluate simple (non-causal) indeterminism and (non-causal) reasons explanations effectively, it is crucial to situate them in the broader debate on the free will problem. First, it is worth noting simple determinism aims to establish libertarianism – a position which affirms both the incompatibility thesis and the free will thesis, but deny the determinism thesis. Even if we concede the incompatibility thesis to the simple determinists, the

incompatibility thesis alone is insufficient to establish the free will thesis. Additional requirements are necessary for the free will thesis – the most fundamental of which is the possibility of action as opposed to mere occurrences and happenings. This fundamental requirement is consistent with both causal determinism and causal indeterminism. A further requirement is the possibility of free actions. That is, given exactly the same past and laws, agents must be able to act otherwise than they actually did. This additional requirement is consistent with causal indeterminism but not causal determinism.

For Ginet and other simple indeterminists, actions are free if they are (1) undetermined, (2) possess actish phenomenal qualities, and (3) done for a reason or purpose. I would argue that not only are these three conditions, as Ginet conceives them, insufficient for the freedom of action, they are also jointly insufficient for action itself. The insufficiency of the first condition is well-known – if determinism (the thesis that every event has a sufficient causal condition) is incompatible with free action, then indeterminism (the thesis that some events like actions are uncaused) is incompatible with free action too. The reason is that most libertarians (other than the simple indeterminists) would want free actions to be caused by the agents themselves so that they can be morally responsible for them. Saying that free actions are uncaused would mean that agents play no role in causing them and hence are not morally responsible for them. The insufficiency of the second condition is this: if we can be mistaken about what seems to us in general, how can having actish phenomenal properties guarantee (in part at least) that we acted freely (or even acted at all)? These actish phenomenal properties may well be mere by-products or effects of certain action-like occurrences and happenings in us. Put simply, actish phenomenal properties of our free action may just be epiphenomenal properties. The insufficiency of the third condition is brought about by its compatibility with causal determinism – for according to the causal determinists, our reasons

and purposes are the content of our intentional states, and our intentional states with reasons and purposes as content themselves have sufficient causal conditions including both factors internal and external to us. If this is the case, then the fact that our actions are done for our reasons and purposes cannot be used to show that they are free and uncaused. And if reasons and purposes are not causes of action at all, as insisted by Ginet and the simple indeterminists, do they then have anything to do with moving our bodies (a fundamental requirement for action)? Of course, Ginet defends the non-causalist account of reasons and purposes against the causalist one. If his defence is successful, then he may be able to show that actions that are done for reasons and purposes are free and uncaused. Let's explore in further detail.

Davidson's causalist account of reasons explanation requires a causal connection between reasons and actions. It is required because 'S has a reason R to do A' does not entail 'S did A because of reason R', and the former entails the latter only when a causal connection 'reason R causes S to do A' is added. Ginet rejects the causal connection because he believes that we can formulate a sufficient truth condition for reasons explanations without entailing that the explanans (reasons) caused the explanadum (actions). For him, the sufficient truth condition for 'S did A in order (with an intent) to B' is something like 'Concurrently with S's doing A, she intends that by doing A she would B'. B can read as a reason or purpose for S to do A. Intention and action are related in two ways – internal and external. The internal relation between intention and action consists of their intrinsic properties, and the external relation between intention and action consists of the intention's direct reference of action as its content. None of these relations imply a causal relation. Even if we concede that Ginet succeeds in providing sufficient truth conditions for reasons explanations without appealing to causal connections, it does not show how this can help the libertarian course in establishing

the free will thesis. As discussed, the free will thesis requires that not only action, but also that free action, is possible. Broadly speaking, action means voluntary bodily movement. Some may object that some actions are purely mental as they do not involve bodily movements. But here, I take action to include neural activities even if the body is still. Whether 'S does A for reason R' or 'S does A in order to B', S must be able to voluntarily move his body in the first place. At this point, the voluntary movement of S's body is consistent with both causal determinism (R or B is causally sufficient for S to do A) and causal indeterminism (R or B inclines but not necessitate S to do A). Notice that both these positions agree that the voluntary movement of S's body is a causal notion, but they disagree over whether or not R or B is the sufficient cause of S's voluntary movement. For simple indeterminism however, R or B is not a causal notion and something other than R and B are the sufficient causal conditions of the voluntary movement of S. Ginet agrees because reasons explanations provide sufficient truth conditions and not sufficient causal conditions for action. It follows then that R or B does not make action possible, something other than R or B does. And if R & B does not make action possible, then a fortiori, it does not make free action possible. Hence, the non-causalist account of reasons explanation does not establish the free will thesis.

Ginet's non-causalist argument against conceiving the intention-action connection causally is this: (a) direct reference requires that what is referred to – the action – causes what refers to it – the intention, and (b) causes must always precede their effects, then (c) there will be a small period at the beginning of the action during which the concurrent directly referring intention will not be in place, a gap during which the agent will have no intention about the action already begun, during which it will not be the case that she intends that action. (c) is implied by (a) and (b) and is a very unattractive consequence. To reject (c), we have to reject either

(a) or (b) or both. Both (a) and (b) seem reasonable if the intention-action connection is causal. For (a), it is difficult to accept that direct reference does not involve the direct referent's entering into causing that which does the referring. For (b), it is difficult to accept that the possibility of an efficient cause that occurs simultaneously with (if not later than) the effect. Ginet's suggests that (c) can be rejected if the intention-action connection in (a) is non-causal.

My reply is that the causalists can accept (b), deny (a), and hence deny (c). I believe that the causalists can agree with Ginet that it is difficult to accept that direct reference does not involve the direct referent's entering into causing that which does the referring, and yet claim that Ginet is mistaken about what the direct referent is. (c) follows from the conjunction of (a) and (b) when the direct referent that figures as the content or object of intention is the actual action itself. Causalists can reject this 'action as content or object of intention' approach by insisting that only the 'representation of action', and not the 'actual action' itself, can be the content or object of the agent's intentional state. And they can accept that the content or object of the agent's intentional state is simply 'to perform the action for a reason or purpose' instead of the 'actual action' itself. That is, the content or object of S's intention refers to 'to do A for R' or 'to do A in order to B' rather than the actual A itself. At this point, the non-causalists can object that R or B is a teleological notion. R or B is a future end that S aims to achieve by performing A. So, a teleological cause like R or B would always succeed rather than precede A. Causalists can respond by saying that 'R or B (a future end that S aims to achieve)' refers to the content or object of S's desire state, and 'A (an action) leads to (or increases the probability of achieving) R or B (a future end that S aims to achieve)' refers to the content or object of S's belief state, and these may result in S's intentional state with the content or object 'to do A for R' or 'to do A in order to B'. These

can be seen as antecedent causal conditions. They need not be some future end that possesses backward causal powers. Read this way, the conjunction of (a) and (b) does not lead to (c).

My criticism of non-causalism so far may not show that causalism is true but it does show that causalism makes action (if not free action) possible by allowing R or B, as the content or object of the agent's intentional states, to cause the voluntary movement of the body where non-causalism does not. Causalism can be qualified in three ways to make it more plausible. First, the causalists do not need to say that reasons are the sole causes of a particular choice or action, and they only need to say that reasons are among the many possible causes of choices or actions. Second, causalists can remain neutral between the relation of intentional states (with reasons as content) and neural states, be it dualism, realisation, supervenience, or identity. Third, the causalist account is consistent with both causal determinism and causal indeterminism. If causal determinism is true, then the agent's having intentional states with reasons as content (together with other possible causes) are causally sufficient for choices or actions. In responding to Ginet's argument against deterministic causalism, causalists like Davidson need not identify reasons as the sole cause of action. And if causal indeterminism is true, then the agent's having intentional states with reasons as content (together with other causes) contribute causally to but are not causally sufficient for choices or actions. To borrow Leibniz's terms, indeterministic causalists can say that reasons and purposes (as content of the agent's intentional states) 'incline without necessitating' choices or actions. In answering Ginet's argument against indeterministic causalism, at least indeterministic causalism provides a necessary if insufficient causal condition for the possibility of choice or action whereas simple indeterminism provides none. By providing antecedent sufficient conditions (if causal determinism is true) or by providing simply antecedent conditions (if causal indeterminism is true) for choice and action, causalism is able to at least secure

ontological conditions for choice and action (or perhaps the free will thesis if other adequate premises are added). By providing merely sufficient truth conditions, non-causalism may indeed succeed in justifying reasons explanations, but they are at most epistemological conditions for the descriptions of action with no ontological import. Only ontological (causal) conditions and not epistemological (truth) conditions, whether sufficient or not, make actions (if not free actions) possible. So far, it seems that causalism, rather than non-causalism, makes action possible. And it also seems that indeterministic causalism, rather than non-causalism, makes free action possible.

For causalism, R or B as the content of enduring states of S like desires, beliefs and intentions is the cause of action. One motivation behind non-causalism is that these enduring states of S with R or B as their content have antecedent causal conditions. If they have antecedent causal conditions, then they are not free. And if these enduring states of S with R or B as content are not free, then A that results from them cannot be free. For deterministic causalism, these antecedent causal conditions are sufficient, and the threat to free action seems obvious. Indeterministic causalism does not seem to solve the problem because of causal gaps. For if S's enduring states with R or B as content incline but not necessitate A, then there is always a chance that S's enduring states with R or B as content will lead to not-A. This causal gap or chance factor is beyond S's control. And as I have argued above, non-causalism does not help matters by merely providing sufficient truth conditions for reasons explanations of action. But if S's enduring states with R or B as content are not the causes of action, then what other options are there? The first option is the behaviourist (purely external) stimulus-response model. This would reduce actions to mere occurrences or happenings and hence unacceptable. The second option held by Timothy O'Connor (O'Connor 2009) retains R or B as non-causal reasons explanation of action, and adds that S

is the sole (purely internal) cause of action. The third option held by Randolph Clarke (Clarke 2009) retains R or B as partial cause, and adds that S complements R or B as the cause of action. The second and third positions are known as agent causal accounts which I will not be discussing here.

For now, I would like to outline a causalist option I favour. First, I adopt the well-accepted assumption that causal relatas are events. Second, I adopt Jaegwon Kim's definition of an event as object (or a related set of objects) exemplifying a property (or properties) at a certain time (or duration of time). Third, I conceive of 'S exemplifying psychological states with certain contents' at an earlier time t0 (or t0 to t1)' as a single cause-event that causes 'S performing an action at a later time t2 (or t2 to t3)' as a single effect-event. Some clarifications are needed here. The cause-event consists of an agent exemplifying psychological states with certain contents at a certain time (or duration of time). Reasons or purposes are the contents of psychological states. And I prefer to think of reasons or purposes as some facts about the situation that an agent finds himself in and responds to. An agent may exemplify psychological states with certain contents at certain times (or durations of time) and sustain them until the agent's action fulfils the specified content. That is, agents exemplifying psychological states with certain contents are enduring states that are extended in time. The agent, his psychological states, and their contents are indispensible constituents of the cause-event. Taken apart, each of the constituents can never be the cause-event. Without his psychological states and their contents, the agent alone lacks the motivational structure required for action. Without his affective, cognitive, and conative capacities, the agent's psychological states and their contents, as well as action can never be exemplified. Without the contents of his psychological states, the agent is acting for no reasons or purposes and such cases can hardly be called action. Moreover, desires, beliefs, and

intentions are distinct psychological states. An intentional state involves a commitment towards action after deliberation while a desire or belief state does not, but a desire state with content 'B' and a belief state with content 'A leads to B' are antecedent causal conditions for an intentional state with content 'to do A in order to B'. Put another way, intentions are post-deliberation or considered wants and desires are pre-deliberation or unconsidered wants, and deliberation or consideration involves beliefs about the wants. The effect-event consists of an agent performing action A. My definition of action involves voluntary bodily movement but exclude its effects. That is, action involving voluntary bodily movement may or may not succeed in achieving its intended effects. Fourth, 'S exemplifying psychological states with certain contents' at an earlier time t0 (or t0 to t1)' may be part of a set of sufficient causal condition for 'S performing an action at a later time t2 (or t2 to t3)'. As discussed, this causalist option has three features as discussed above: (a) 'S exemplifying psychological states with certain contents' at an earlier time does not have to be the sole cause of 'S performing an action at a later time', (b) it remains neutral to the relation between psychological states and neural states, be it dualism, realisation, supervenience, or identity, and (c) it remains neutral between causal determinism and causal indeterminism.

The motivation behind my causalist option is not to establish freedom of choice or action. Rather, it is to make conditioned choice or action possible. If causal determinism is true, then there are sufficient causal conditions for our choices and actions. In this case, we do not have full control over our choices and actions. But it does not seem true to say that we have no control at all because our affective, cognitive, and conative capacities, together with our psychological states and reasons as their contents (though they themselves have sufficient causal conditions) are factored in as co-determinants of our choices and actions. And if causal indeterminism is true, then there are insufficient causal conditions for our choices and

actions. In this case, we do not have full control over our choices and actions too. But again, it does not seem true to say that we have no control at all because at least our affective, cognitive, and conative capacities, together with our psychological states and reasons as their contents are causally relevant to our choices and actions, despite the presence of causal gaps or chance factors.

**Three Non-Causalist Objections and Causalist Replies**

There are several possible objections the non-causalists can raise against the causalist account. First, non-causalists can claim that not all actions involve voluntary bodily movements for there are purely mental actions. Hence, voluntary bodily movement is not a requirement for action. But in attempting to establish the possibility of free action and our moral responsibility for them, it seems obvious that we are more interested in actions that involve voluntary bodily movements that can affect the world around us, much more so than purely mental actions. Given our best scientific theories, voluntary bodily movements have causes, whether deterministic or not. More contentious is the claim that reasons cause voluntary bodily movements. As discussed above, my proposal is that reasons, as contents of psychological states had by the agent, cause voluntary bodily movements only indirectly. So far, I am not committed to any account of how the agent having enduring states with content is related to his neural states, just that the former has to figure as part of the causal explanation for voluntary bodily movement, in order to establish the possibility of free actions and our moral responsibility for them. In insisting that actions involving voluntary bodily movements are uncaused, non-causalist accounts are at odds with both our best scientific theories as well as the libertarians' attempt to establish the possibility of free action involving voluntary bodily movement and the agent's moral responsibility for them. Second,

non-causalists can claim that if intention (with a reason as content) causes action, then intention and action cannot be logically connected. They can contend that intention and action are logically connected, as it seems that intention forms a necessary part of action or that there seems to be no action without intention. Hence, the relationship between intention (with a reason as content) and action cannot be a causal one. For if intention is 'part of' action, then it cannot be the 'cause of' action at the same time. This is a version of the well-known logical connection argument advanced against causalism. Davidson famously replied that the relatas of causation are events no matter how they are described, and they are not event-descriptions. Non-causalists may describe intention (with a reason as content) as part of action rather than its cause, but this does not imply that intention-action as a single event has no cause (the point of Davidson's Bridge example). Even if causalists concede that the agent exemplifying an intention state with reason as content is 'part of' rather than the 'cause of' an action, they can still insist that the agent exemplifying desire-belief state (with reasons as content) is the cause of his intention-action. It is implausible for non-causalists to hold that the agent's desire-belief state (with reasons as content) is 'part of' rather than the 'cause of' his actions. And even if non-causalists insist that the agent's desire-belief state (with reasons as content) is 'part of' rather than the 'cause of' actions, this does not imply that a desire-belief-intention-action as a single event has no causes. As argued above, actions involving voluntary bodily movements are directly relevant to the establishment of free actions and our moral responsibility for them. Here, non-causalists can either hold that voluntary bodily movements have no causes at all or that the causes of voluntary bodily movements are entirely physical and they do not include our desire-belief-intention states at all. The former claim seems implausible given our best scientific theories and the latter claim seems unpalatable to the establishment of free actions and our moral responsibility for them.

Third, non-causalists can claim that actions can have causes and effects deviant from the agent's intentions (with reasons as content). This is commonly known as the deviant causal chains objection against causalism. Actions that have causes deviant from the agent's intentions (with reasons as content) is named primary deviance. And actions that have effects deviant from its intentions (with reasons as intent) is named secondary deviance. A standard reply by causalists to cases of secondary deviance is that they are unintended effects or accidents beyond the control of agents and hence are not actions. My own reply to secondary deviance is to exclude the effects of action from the definition of action itself, or hold that actions do not necessary entail their effects. Secondary deviance objections conflate action with the effects of action. When S performs A in order to B, it remains possible that S did perform A but fails to achieve B. Moreover, it seems implausible to include the effects of action in the definition of action, or hold that an action necessarily entails its effects. On this view, when S performs A in order to B, it is impossible that S did perform A but fails to achieve B. The consequence of this view then is that there is no such thing as a failed action, which is rather counter-intuitive. Primary deviance remains the more serious objection against causalism. A standard reply by causalists to cases of primary deviance is to specify that if our intentional states (with reasons as content) are among the causes of our actions, then they are caused in the 'right' or 'non-deviant' ways. One concession the causalists can grant is that we may not know or may not be conscious of the 'real' content of our intentional states that causes our action states, but this does not mean that some intentional states (with reasons as content) are not the causes of our actions. Another concession the causalists can grant is that intentional states (with reasons as content) may not be the sole causes (or may not be the sufficient causal condition) of actions but they may nonetheless be among the causes. That is, causalists can accept that all actions imply intentions but not all intentions imply actions.

Furthermore, I shall argue that deviant causal chains can be even more devastating to the non-causalists. If actions involve voluntary bodily movements, and if our voluntary bodily movements are effects of deviant causal chains such that our intentional states (with reasons as content) are not among the causes, then it seems that these voluntary bodily movements are beyond the control of agents, and hence they can hardly be called actions. In comparison, causalists can still appeal to actions involving voluntary bodily movements having non-deviant causal chains (or being caused in the right way) to eliminate the possibility of deviant causal chains. But this option, however, is not open to the non-causalists who insist that the agent having intentional states (with reasons as content) are not among the causes of actions states involving voluntary bodily movements. That is, non-causalists have no means to avoid the effects of deviant causal chains on voluntary bodily movements.

**Teleological Explanation as Non-Causal Explanation of Action**

As mentioned above, there are two ways to explain events for the non-causalists – in terms of causes and in terms or reasons and purposes. When explained in terms of causes, events are happenings and occurrences. And when explained in terms of reasons and purposes, events are actions. Events that are explained in terms of causes are called causal explanations and actions that are explained in terms of reasons and purposes are called teleological explanations. Non-causalists typically hold that actions can only be explained in terms of reasons and purposes and they cannot be explained in terms of causes. In other words, actions are subject to teleological explanations and not causal explanations. I take this to be a serious objection against causalism and I shall briefly discuss this objection as recently expressed by Scott Sehon (2010), Stewart Goetz (Aguilar and Bukareff 2009), Hugh McCann

(Aguilar and Bukareff 2009), and George Wilson (Aguilar and Bukareff 2009, Wilson 2007) as well as my replies to them.

Scott Sehon holds that teleological explanations of an action would mention the agent, the agent's behaviour, and the state of affairs – which is the end, goal, purpose, or reason – towards which the agent's behaviour was directed. On his non-causalist view, the fact that an agent is directing his behaviour to some states of affairs is not grounded in, or reducible to some further facts, like causal factors in terms of psychological states. And since teleological facts do not reduce to causal facts, definitive truth conditions for teleological explanations cannot be given in causal terms. For Sehon, rationality and value seem be the two truth conditions of teleological explanations of action. While the rationality condition refers the degree to which the explanation makes the behaviour appropriate for achieving a goal, the value condition refers to the degree to which the goal in question is of value. And he seems to claim that these conditions cannot be accommodated by causal explanations of action, and hence reduction of teleological explanation to causal explanation is not possible.

Causalists can contend that the rationality and value conditions can be accommodated by causal explanations. Sehon seems to conflate causalism with reductionism but causalism need not entail reductionism. He would be right if he claims that the rationality and value conditions may not be satisfied by a causal explanation of action, but he would be mistaken if he claims that rationality and value conditions can never be satisfied by a causal explanation of action. Perhaps distinguishing between motivating reasons and normative reasons could clarify matters here. Motivating reasons (reasons why an agent does something) explain actions and normative reasons (reasons for which an agent does something) justify actions (Smith 1994, pp.94-8). Normative reasons reflect the rationality and value conditions. Here, causalists can remain neutral between instrumental and intrinsic rationality, and between

prudential and moral value. Even if motivating (explanatory) reasons may not be normative (justificatory) reasons, it is a mistake to claim that normative (justificatory) reasons can never be motivating (explanatory) reasons. When an agent does act in accordance with rationality and value conditions, his motivating reasons (explanatory = reasons why an agent does something) is also his normative reasons (justificatory = reasons for which an agent does something). And when an agent does not act in accordance with rationality and value conditions, his motivating reasons and normative reasons come apart. According to causal explanations of action, only motivating reasons, whether or not they are also normative reasons, explain what moves the body towards certain behaviour. Normative reasons alone do not for they are just rationalisations and justifications. Notice that "move" is itself a causative verb. Teleological explanations of action, if they are explanations of action at all, must explain what move the body towards certain behaviour by citing normative reasons which are also motivating reasons. And when they do, teleological explanations are also causal explanations. Suppose B is an end, goal, purpose, or reason valued by an agent. The agent values goal B, and knows that behaviour A is appropriate for achieving goal B, decides to execute behaviour A in order to achieve goal B, and hence executes behaviour A. This value-knowledge-decision model of teleological explanation is structurally similar to the desire-belief-intention model of causal explanation. And this causalist view of teleological explanation not only accommodates the agent, his behaviour, the goal towards which the agent's behaviour was directed, as well as rationality and value conditions, but also connects them causally to the movement of the body. Such causal connections are not captured by a non-causalist view of teleological explanations, where rationality and value conditions play no causal roles in moving the body that is essential to the explanation of action or bodily movement. Hence, causal explanations and teleological explanations need not be mutually exclusive. And since the causalist view of teleological explanation of action is more

comprehensive (able to state normative reasons which are also motivating reasons that move the body) than its non-causalist counterpart (able to state normative reasons that do not move the body, but not motivating reasons that move the body), the causalist view of teleological explanation should be preferred.

Stewart Goetz holds that the agent's choice can only be teleologically explained by a reason and it cannot be causally explained. A reason for choosing at act is a purpose, where a purpose is a conceptual entity, which is the content or object of an intentional state, which is in turn about or directed at the future. When an agent chooses to act, he does so in order to accomplish or bring about a purpose in the future. Teleological and causal explanations have different directions of fit. While teleological explanations have a future-to-present direction of fit, causal explanations have a past-to-present direction of fit. Hence, to propose a causal explanation of a choice is fundamentally to misunderstand or misrepresent the correct explanatory direction for fit for a choice.

My proposed causalist view can accept that reason or purpose refers to the content or object of an intentional state, which is possibly a conceptual entity (as mentioned, I prefer to think of a reason or purpose as facts about the situation that an agent finds himself in and responds to) that is about or directed at the future. But the cause in my proposed causalist view does not refer solely to the content or object of an intentional state. Rather, it refers to an event where an agent exemplifies intentional states with an object or content at a certain time (or duration of time). On my view, a reason or purpose does not explain the agent's choice or action unless it figures in the agent's desire or belief states as their object or content, which are the agent's attitude or response towards the reason or purpose. A reason or purpose explains the agent's action only when he adopts a certain attitude or responds in a certain way towards a reason or purpose, which is expressed in terms of his intentional states.

Abstracting the agent's states away from their content in the explanation of action implies the irrelevance of the agent's attitude or response towards his reasons or purposes in the explanation of action. This is implausible because it is possible that a reason or purpose to act may be present but the agent forms no attitude or response towards it and hence fails to act upon it. To say an agent chooses to act upon a reason or purpose B is to say that he desires B, and believes that action A leads to B, and hence intends to do A in order to B. Seen this way, it is not the case that teleological explanations have a future-to-present direction of fit because the reason or purpose has to figure as the object or content of the agent's intentional states (which are antecedent causal conditions) in order to explain the agent's choice or action. Teleological explanation is seen as a form of causal explanation that involves the agent's intentional states, with reasons and purposes as their object or content, which is about or directed at the future. In this sense, teleological explanations can be said to have a past-to-present (or present-to-future) direction of fit. Hence, there is no real conflict between teleological explanations and causal explanations of choice and action, and there is no good reason why choice and action cannot be causally explainable even if they are teleologically explainable at the same time.

Hugh McCann argues that in teleological explanations, the agent's reasons have explanatory force qua reasons, and discusses two important points that follow from this view. First, if it is the content of the agent's desire that explains here, then the explanation would not be causal, at least not in the sense of nomic causation. This is because content is not an event but an abstraction (like Goetz's conceptual entity), and nomic causes have to be events. Yet, the content explains what I did, at least to some extent. And if we accept this, then we must be prepared to accept the legitimacy of (teleological) explanations that are not causal. Second, if teleological explanation proceeds in terms of the mental content of the agent's desire, then it

is inherently dualistic and it is hard to see how teleological explanation retains its explanatory force in a reduction of the mental to the purely physical. Even if it is assumed that our mental events or states are identical with or supervenient on physical events or mental states, it will still be the case that causal physical explanations of voluntary bodily movements are incomplete, unless they are supplemented with teleological (non-causal) mental explanations.

In reply to the first point, I contest the claim that in teleological explanations, the agent's reasons have explanatory force qua reasons alone. As mentioned above, what also matters significantly is the agent's attitude or response towards his reasons expressed in terms of his desires and beliefs. It is not solely the content of the agent's desire that explains here, but rather the agent's having intentional states with certain contents. And the agent having intentional states with certain contents at certain times (or durations of times) can be considered mental events, not abstractions. It may be objected that since the content is sufficient to explain action, why bother to bring in intentional states or even the agent as substance? My reply is that action requires voluntary bodily movement and as argued above, I do not think that content alone (an abstraction and not an event) is sufficient to explain it. On the causalist view I proposed, it is the agent having intentional states with certain contents at certain times (or durations of time) that causes his actions. If this is the case, then teleological explanation can be a form of causal explanation that takes into account the agent's reason and purpose, even if it's disputed whether they can be expressed in a nomic way or not. There are other non-nomic ways of expressing causal relations, such as the difference-making, counterfactual supporting, contributory, or INUS accounts.

In reply to the second point, causalists can agree with McCann that (1) teleological explanations are inherently dualistic , (2) teleological explanations lose their explanatory force in a reduction of the mental to the purely physical, and (3) causal physical explanations

of voluntary bodily movements are incomplete unless they are supplemented with teleological mental explanations. But under the causalist view, (1) it can be assumed that dualistic teleological explanations are causal instead of non-causal, (2) if (causal) dualism is true, then distinct and irreducible mental events and states can cause physical states or events, and if physicalism is true, it can still be the case that the mental events or states are ontologically reducible to physical events or states in a broad sense, even if (causal) mental explanations are not epistemologically reducible to (causal) physical explanations, (3) under either (causal) dualism or physicalism, it can still be true that causal physical explanations are incomplete unless they are supplemented by teleological mental explanations. And as discussed, causalists can remain neutral between on the relationship between the mental with the physical, be it dualism, realisation, supervenience, or identity. Whatever option is chosen, the causalists are in a better position to explain voluntary bodily movement by showing how the agent's having intentional states with reasons as contents makes an ontological difference to his voluntary bodily movement, by integrating teleological and causal explanations. And by isolating teleological from causal explanations of voluntary bodily movement, non-causalism is unable to adequately show how it is possible for the agent to move his body voluntarily for reasons and purposes.

George Wilson argues against the causalist view of teleological explanations, starting with a criticism of Davidson's account. Davidson supposes that teleological explanations of action are themselves just causal explanations of a certain kind. Very roughly, Davidson believes that in giving the reason or purpose of an agent's action, we are in effect saying that the agent had the relevant reason or purpose, and the agent having that reason or purpose was, among other factors, a cause of his action. Wilson objected that this constitutes a strong conceptual claim for which Davidson provides no independent argument. While this objection may be

true, it does not show that an independent argument for this claim is not available. Note again that when we are talking about action, we are also talking about voluntary bodily movement. The motivation for conceiving the agent having the relevant reason or purpose as a cause of his action is to explain how the agent can move his body voluntarily by having the relevant reason or purpose. This is why causalists like Davidson suppose that teleological explanations are just a kind of causal explanation. At first glance, non-causalists like Wilson who insist that teleological explanations are not causal cannot adequately explain how the agent can move his body voluntarily by having the relevant reason or purpose. Wilson then proceeded to elaborate on his own non-causalist account of teleological explanation. He holds that intentional action is under the guidance of the agent, and that it is exercised in an epistemically privileged fashion (from the first-person perspective). In guiding his intentional movements, the agent has knowledge of both the goal of his behaviour as well as the evolving trajectory in the situation. This knowledge cannot be derived from observation of the agent's behaviour or the situation in which it occurs (from a third-person perspective). It is the agent's epistemically privileged (or first-person) guidance of his behaviour towards the goal that grounds teleological explanations of action, and this is not reducible to the non-epistemically privileged (or third-person) causal explanations.

I have two replies to Wilson's account. First, causalists need not reject the epistemically privileged or first-person feature of teleological explanations. Put another way, this epistemically privileged feature or first-person feature is not exclusive to a non-causalist view of teleological explanation. One example is that causalists who are also dualists can hold that intentions or volitions, conceived as irreducible and epistemically privileged mental events or states, cause voluntary bodily movements. Another example is that causalists who are also physicalists of the non-reductive variety can hold that certain physical events or states that

cause voluntary bodily movements simply possess epistemically privileged or first-person phenomenal properties. Whether combined with dualism or the non-reductive form of physicalism, the causalists are in a better position to explain voluntary bodily movement by showing how the agent's having intentional states with reasons as content makes an ontological difference to his voluntary bodily movement, by integrating teleological and causal explanations. Second, basing the agent's control and guidance of his actions on phenomenology without ontology (like Ginet's actish phenomenal properties) is inadequate. For it is always possible that the agent's epistemically privileged or first-person properties are epiphenomenal and that they do not make an ontological difference to his voluntary bodily movements.

Wilson grants that states of wanting, desiring, and believing can cause all sort of involuntary behaviour (or unintentional action), including behaviour which may lead to the agent's desired ends by accident. For this reason, causalists fail to capture the most important feature of voluntary behaviour (or intentional action), where it is the agent who controls and guides the behaviour (or action) towards the desired goal. He quotes from David Velleman's essay 'What Happens When Someone Acts?' to illustrate this point: "In a full-blooded action, an intention is formed by the agent himself, but not by his reasons for acting. Reasons affect his intention by influencing him to form it, but they thus affect his intention by affecting him first. And the agent moves his limbs in execution of his intention; his intention doesn't move his limbs by itself" (Velleman 2000, pp.148).

One obvious question here is why assume that states of wanting, desiring, and believing cause only involuntary behaviour or unintentional action. This is counterintuitive to what we normally mean by the terms voluntary or intentional. Doesn't voluntary behaviour or intentional action by definition follow from the agent's wants, desires, and beliefs, regardless

of whether causal determinism or causal indeterminism is true? Perhaps for Wilson, the terms voluntary or intentional apply only to the agent but not his states or their content like reasons. This is very similar to the agent causal account where the agent alone, excluding his states and their content, counts as the cause of a volition or intention. I believe that what differentiates Wilson's non-causal account from the agent causal account is that the former does not view the agent as a cause, but more like a source of control or guidance. I will not comment on the agent causal account here as my purpose is to address the non-causal account. A further question here is what the agent's control or guidance amounts to if they are not causes? The focus here is how the agent moves his body voluntarily or intentionally. Is an agent's control or guidance at least a condition, whether necessary or sufficient, of his voluntary behaviour or intentional action? If it is, then Wilson's account collapses into the agent causal account. If it is not, then it is not at all clear how the agent's control or guidance makes an ontological difference to his bodily movements.

Another important question pertains to the relationship between the agent, his intention, and his reasons. Wilson endorses Velleman's remark that reasons affect the agent's intention by influencing him to form it. Note that on Wilson's account, 'affect' and 'influence' here are not read causally. Here again, are reasons at least an antecedent condition, whether necessary or sufficient, for an agent to form an intention? If they are, then Wilson's account collapses into some form of causalism. If they are not, then it is not at all clear how reasons make an ontological difference to the agent forming an intention. Wilson concedes that there are causal explanations of the behaviour that constitutes action. Neurophysiological studies of voluntary movement are one important source of causal accounts. There is in principle no reason why these accounts cannot be made detailed and complete. And even if this is the case, teleological explanations and causal neurophysiological explanations of intentional

action need not stand in any exclusionary relation to one another. He believes that these distinct explanations account for the same events (the agent's actions) but answer different questions about them.

I agree with Wilson that teleological and causal explanations of action are not mutually exclusive but I disagree with him in (1) isolating both types of explanations and (2) holding that each type of explanation can be in principle sufficient and complete as an explanation of action. In explaining action involving voluntary bodily movement, we are interested in what makes the body move voluntarily. On the one hand, if the causal neurophysiological explanation is both sufficient and complete for voluntary bodily movement on its own, then it makes the non-causal teleological explanation redundant. And on the other hand, it is hard to see how a non-causal teleological explanation can be both a sufficient and complete explanation of voluntary bodily movement on its own, without a causal neurophysiological explanation. For how can the agent's reason move his body if it is not conceived at least as an antecedent condition, whether necessary or sufficient, for voluntary bodily movement. Of course, non-causalists prefer to use terms like the agent's reasons 'affecting', 'guiding', or 'influencing' rather than 'causing' voluntary bodily movement. Another important question is whether affecting, guiding, or influencing contributes to moving the body. If they do, then why not say that the agent's reasons cause his voluntary bodily movement. And if they do not, then they do not explain voluntary bodily movement at all.

In his paper 'Action' (Wilson 2007), Wilson discusses the Principle of Explanatory Exclusion and its implications for causalism and non-causalism with reference to Jaegwon Kim's paper 'Mechanism, Purpose, and Explanatory Exclusion' (Kim 1997). The Principle states that no event or phenomenon can be given more than one 'independent' and 'complete' explanation. Influenced by Davidson, many philosophers reject more than bridge laws between reason

states and action states. They believe, more generally, that there are no laws that connect any mental states, events, and processes with *any* physical states, events, and processes. As a consequence, vernacular psychology is not strictly reducible to the neural sciences. This means that reason explanations of action and corresponding neurophysiological explanations of action are 'independent' of one another. In short, irreducibility implies independency. Moreover, both reasons explanations and neurological explanations of action are assumed to be 'complete', with each claiming to provide sufficient conditions of action. This implies the overdetermination of action by reasons explanations and neurophysiological explanations and one of them is redundant. Since we are not likely to give up our best and most worked-out scientific accounts, it is our commonsense accounts that appear to be threatened. Wilson believes that causalism is more vulnerable to the Principle than non-causalism, perhaps because under causalism (and not non-causalism), both the reason causes and neural causes compete to make ontological differences to action. And if Explanatory Exclusion applies to causal reason explanations of action, then there is a strong incentive to search for a workable philosophical account of non-causal reason explanations. He argues by way of analogy that just as certain functional explanations in biology may not reduce to, but also certainly do not compete with, related causal explanations in molecular biology, so also non-causal reason explanations could be expected to co-exist with causal neurological explanations of the causes of actions.

Replying to Wilson's argument from analogy, it is still controversial whether functional explanations can or cannot be causal explanations. Examples of functional explanations in biology are 'plants contain chlorophyll so that they can photosynthesise' and 'polar bears are white so that they can camouflage'. My preferred view, as held by most contemporary philosophers of science, is that functional explanations, despite appearances, are really forms

of causal explanations. Their references to future effects are only apparent and they really refer to past causes (Papineau 1997, pp.179). I prefer to state this view as functional explanations can be understood as causal explanations. Using the above examples to understand functional explanations causally, 'plants photosynthesise because they contain chlorophyll' and 'polar bears can camouflage because they are white'. Thus, the functional explanation of the polar bears' colour refers to the fact that their past camouflaging led to the natural selection of their whiteness, and not the fact that they may be camouflaged in the future. If this view is correct, then Wilson's analogy is false. This is because functional explanations in biology and causal explanations in molecular biology are both causal explanations whereas non-causal reasons explanations and causal neurophysiological explanation are not both causal explanations.

Replying to the Principle in general, note that the requirement for Explanatory Exclusion is the conjunction of 'independence' and 'completeness' of more than one explanation for the same event of phenomenon. If it cannot be conclusively shown that either one or both of the conjunct hold, then the Principle does not apply. Kim has made a few suggestions on how this can be done (Kim 1997, pp.269-70). On the independence requirement, causalists may adopt the following options on the relationship between the mental and the physical. First, assuming that there are bridge laws or law-like relations between them, they are identical and are nomic equivalents. Second, whether there are bridge laws or law-like relations between them, the mental are supervenient on or realised by the physical without being reduced. Third, whether there are bridge laws or law-like relations between them, the mental is distinct from the physical and they can act on one another. In the case of identity, there is reduction and there is dependency. In the case of supervenience and realisation, there is no reduction but there is still dependency in some sense. And in the case of dualism, there is no reduction

and there is interdependency. In all three cases, there is either dependency or interdependency and hence the independence requirement is not met. The completeness requirement is reliant on the independence requirement. If there is independency between two or more explanations, then each explanation can either be complete or incomplete. If it is complete, then it is a case of overdetermination and redundancy. If it is incomplete, then it is a case of co-determination but the relation between the co-determinants remains a mystery. And if there is dependency between two or more explanations, then each explanation is incomplete and the completeness requirement is not met. There is no overdetermination, redundancy, or mystery either. All three cases are consistent with the claim that both causal reasons explanations and neurophysiological explanations are dependent and incomplete and hence they are not subject to the Principle of Explanatory Exclusion.

The plausibility of Wilson's claim that non-causal reason explanations may not reduce to, and do not compete with causal neurophysiological explanations and both could be expected to co-exist as explanations of action depends on whether the event or phenomenon to be explained are distinct. The more distinct they are, the stronger Wilson's claim is and vice versa. Non-causalists may want to distinguish between purposive action and voluntary bodily movement and insists that non-causal reasons explanations explain the former whereas causal neurophysiological explanations explain the latter. Kim claims that although the explanadum of both explanations are not equivalent or synonymous, there is an evident sense in which they describe one the same event or concrete happening (Kim 1997, pp.260-1) and I agree. Even if it is argued that purposive actions are not identical with voluntary bodily movements, they do involve voluntary bodily movements. An explanation that claims to explain a particular purposive action has to explain the voluntary bodily movement involved in that purposive action. So, purposive action and voluntary bodily movement can be

conceived as one event or phenomenon under different descriptions. If this is true, and if Wilson insists that both non-causal reasons explanation and causal neurophysiological explanation are both 'independent' and 'complete', then by the Principle of Explanatory Exclusion, one of them has to be given up. And since we are not likely to give up our best and most worked-out scientific accounts, non-causal reasons explanation has to go. In reply, non-causalists can argue that both non-causal reasons explanations and causal neurological explanations are both 'independent' but 'incomplete' to avoid Explanatory Exclusion. I do not think this reply succeeds. Even if causal neurological explanations do not provide sufficient causal conditions of action, what do non-causal reasons explanations add? Non-causal reasons make no difference to, do not counterfactually support, do not contribute towards, and are not INUS conditions of action involving voluntary bodily movement. To avoid Explanatory Exclusion, non-causal reasons explanations are best treated as mere rationalisations and justifications of actions, not real explanations. And if reasons do figure in explanations of action, then they have to make a difference, counterfactually support, contribute towards, or are INUS conditions of action. In short, reasons have to figure causally.

**Teleological Explanation as Causal Explanation of Action**

In view of these difficulties, I propose that a more plausible causalist alternative lies in (1) integrating rather than isolating both types of explanation and (2) holding that neither type of explanation can be in principle complete or sufficient on its own for the explanation of action. For this proposal to work, I have to reject three crucial non-causalist assumptions which I find unwarranted. First, I have to reject the assumption that causal explanations of voluntary bodily movement can only be expressed in neurophysiological terms. Once this assumption

is rejected, we can express the agent having an intentional state with reason as content (whether the intentional state with reason as content is distinct from a neural state in the dualistic sense, or whether it is identical with, supervenient on, or realised by a neural state) in causal terms. This allows us to integrate both types of explanations or more pointedly, treat teleological explanation as a form of causal explanation. The motivation behind the integration (rather than isolation) of both explanations is to show that the agent and his enduring states with reasons as content make an ontological difference to his voluntary bodily movement (and not just an epistemological difference with no ontological import). Second, I have to reject the assumption that causal explanations must involve strict causal laws. That is, if the causal relatas cannot be subsumed under strict causal laws, then there cannot be any causal relation between them. There may possibly be non-strict (ceteris paribus) laws governing human action but my proposed account is not committed to any laws governing human action, be it strict or non-strict. They may be singularist causal explanations. Moreover, my proposed causalist alternative is not committed to either realism or anti-realism pertaining to causation. All it needs is a difference-making, counterfactual supporting, contributory, or INUS account of causation. Third, I have to reject the assumption that the only relevant feature in reasons explanation or teleological explanation is reasons as mental contents. On my proposed causalist alternative, the agent, his psychological states, and their contents are indispensible constituents of reasons explanations or teleological explanations. What explains action (voluntary bodily movement) is the agent exemplifying intentional states with reasons as contents, conceived as events. Suppose B is an end, goal, purpose, or reason. The agent desires B, believes that an action A leads to (or increases the probability of achieving) B, and intends to do A in order B, and hence does A. Reasons as mental contents alone explain little without the intentional states as well as the agent himself. The intentional states exemplified by the agent are necessary to explain his

attitude or response towards the reason and the intentional state had by the agent is necessary to explain his commitment towards action (voluntary bodily movement). And of course, intentional states do not float free; they must be had by some agent, who possesses the necessary affective, cognitive, and conative capacities. If this proposal is plausible, then it seems that the causalist notion of reasons explanation or teleological explanation proves more adequate in accounting for action involving voluntary bodily movement than the non-causalist alternative.

My causalist account as discussed is not without problems. Even if it is granted that intentions and actions require causes, the causes need not be psychological states with reasons as their content. As it is my view that psychological states with reasons as their content can have antecedent causal conditions, whether deterministic or indeterministic, it seems like agents are passive (being acted upon by events) rather than active (acting upon events). It then follows that my causalist approach does not seem to support the ascription of moral responsibility in a robust sense. To support the ascription of moral responsibililty in a robust sense, the causes have to be the agents as substances rather than their exemplification of psychological states with reasons as their content. Only agents as substances which cause their intentions and actions can be conceived actively (acting upon events) rather than passively (being acted upon by events). Moreover, the agents cannot have antecedent causal conditions in virtue of being substances. This alternative causalist account is known as agent causation. To this, my brief reply is that the conception of an agent as a substance which is an uncaused causer alone is unintelligible. When asked why an agent as substance intends and acts the way he does, the answer is nothing because substances by definition cannot have antecedent causal conditions. As a rejoinder, the agent causalists may claim that an agent's intentions and actions can be made intelligible by considering non-causal reasons. The

problem with this rejoinder is that since an agent as substance is the sole cause of his intentions and actions, reasons are not causally relevant as they make no difference to, do not counterfactually support, do not contribute towards, or are not INUS conditions of his intentions and actions.

**References:**


Aguilar, Jesus H., Buckareff, Andrei A. ed. (2009) *Philosophy of Action* (Automatic Press)

Clarke, Randolph (2009) from 'Agent Causation and Event Causation in the Production of Free Action' in *Free Will* ed. Derk Pereboom (Hackett)

Clarke, Randolph (2002) 'Libertarian Views: Critical Survey of Non-Causal and Event-Causal Accounts of Free Agency' in *The Oxford Handbook of Free Will* ed. Robert Kane (Oxford University Press)

Davidson, Donald (1980) *Essays on Actions and Events* (Oxford University Press)

Davis, Wayne A. (2010) 'The Causal Theory of Action' in *A Companion to the Philosophy of Action* ed. Timothy O'Connor and Constantine Sandis (Blackwell)

Ginet, Carl (2002) 'Reasons Explanations of Action: Causalist versus Non-causalist Accounts' in *The Oxford Handbook of Free Will* ed. Robert Kane (Oxford University Press)

Ginet, Carl (1990) *On Action* (Cambridge University Press)

Kane, Robert (2005) *A Contemporary Introduction to Free Will* (Oxford University Press)

Kim, Jaegwon (1997) 'Mechanism, Purpose, and Explanatory Exclusion' in *The Philosophy of Action* ed. Alfred Mele (Oxford University Press)

Mele, Alfred ed. (1997) 'Introduction' in *The Philosophy of Action* (Oxford University Press)

O'Connor, Timothy (2010) 'Reasons and Causes' in *A Companion to the Philosophy of Action* ed. Timothy O'Connor and Constantine Sandis (Blackwell)

O'Connor, Timothy (2009) 'Agent-Causal Power' in *Free Will* ed. Derk Pereboom (Hackett)

Papineau, David (1997) 'Methodology: The Elements of the Philosophy of Science' in *Philosophy: A Guide through the Subject* ed. A.C. Grayling (Oxford University Press)

Sehon, Scott (2010) 'Teleological Explanation' in *A Companion to the Philosophy of Action* ed. Timothy O'Connor and Constantine Sandis (Blackwell)

Smith, Michael (1994) *The Moral Problem* (Blackwell)

Stout, Rowland (2010) 'Deviant Causal Chains' in *A Companion to the Philosophy of Action* ed. Timothy O'Connor and Constantine Sandis (Blackwell)

Velleman, J.D. (2000) *The Possibility of Practical Reason* (Oxford University Press)

Wilson, George (2007) 'Action', *Stanford Encyclopedia of Philosophy* (SEP) http://plato.stanford.edu/entries/action

## CHAPTER 3: AGENT CAUSATION AS EVENT CAUSATION

### Introduction

In this essay, I seek to critically evaluate agent causation as separate and distinct from event causation and to defend the claim that agent causation is a form of event causation within the context of the free will debate. On the critical evaluation of agent causal accounts as separate and distinct from event causal ones, I refer to the works of Timothy O'Connor, Randolph Clarke, and E.J. Lowe. First, I will discuss O'Connor's earlier view with reference to his paper 'Agent Causation', where he combines agent-causes with non-causal reasons in the explanation of action. Second, I will discuss O'Connor's later view with reference to his paper 'Libertarian Views: Dualist and Agent-Causal Theories', where he suggests that an agent's recognition of reasons to act induces or elevates an objective propensity to initiate behaviour, but maintains that embracing this causal-propensity account of the relative strength of reasons need not lead one to abandon the non-causal relationship between actions and the reasons that explain them. Third, I will discuss Clarke's view with reference to his paper 'Agent Causation and Event Causation in the Production of Free Action', where he combines agent-causes and reason-causes in the explanation of action. Fourth, I will discuss E.J. Lowe's arguments against the reducibility of agent causation to event causation as well as his arguments for the reducibility of event causation to agent causation in chapter six of his book ''Personal Agency: The Metaphysics of Mind and Action'.

My main objections to agent causation as separate and distinct from event causation are based on the problematic relationship between the agent (solely as substance) and his reasons (whether conceived non-causally or co-causally) in the explanation of action, as well as the

inexplicability, unintelligibility, and implausibility (if not impossibility) of conceiving the agent (solely as substance) as an irreducible sole or co-cause of action. And given this objection, I propose to reject the notion of agent-cause as substance-cause and accept that agent causation is fundamentally event causation involving the agent. My defence of the claim that agent causation as event causation involving the agent is premised on the claim that the most plausible way to causally integrate the agent and his reasons is to treat the agent exemplifying certain psychological capacities and intentional states with reasons as their content or object at certain times (or durations of time) as a single cause and explanation of action, rather than to conceive the agent as the sole cause of action, or to conceive the agent and his reasons as separate and distinct co-causes of action.

**The Motivation behind Agent Causation**

The dilemma of free will is that if actions are caused deterministically, then they are not free; and if they are not caused deterministically, then they are not free either because then they happen by chance and are not up to the agent. What makes agent causation attractive is that it seems to resolve this dilemma by asserting that because actions are self-caused and self-determined, they are entirely up to the agent and hence are neither determined (by the self) nor happened by chance. And why this supposed feature makes agent causation attractive is its implication for moral responsibility. According to the agent-causalists, causal determinism and causal indeterminism are incompatible with the freedom of choice and action because it is not entirely up to the agent. In other words, both determinism and indeterminism do not cohere with the attribution of ultimate moral responsibility. Attribution of ultimate moral responsibility requires absolute freedom of choice and action, which in turn requires that it is entirely up to the agent. So, only agent causation can secure absolute

freedom of choice and action by postulating that it is entirely up to us, which in turn secures the attribution of ultimate moral responsibility by securing an entirely self-causing or self-determining absolute freedom of choice and action. Put another way, only agent causation coheres with the attribution of absolute freedom of choice and action as well as ultimate moral responsibility.

**Event Causation versus Agent Causation**

Proponents of event causation need not commit to either the reductive (non-realist) or non-reductive (realist) causal relation between events. My preferred accounts of causation are causes as difference-makers, counterfactual supporters, contributors, and what J.L Mackie calls Insufficient but Necessary parts of an Unnecessary but Sufficient Condition (INUS) condition (Mackie 1993). One influential account of events by Jaegwon Kim defines an event as an object (or a certain group of objects) exemplifying a particular property (or properties) at a certain time (Kim 1975). In event causation, to say that one event causes another event is to say that an object x exemplifying a property p at time t0 causes an object y exemplifying property q at time t1. In short, event (x, p, t0) causes event (y, q, t1). One limitation of Kim's account is that it fails to capture the fact that events can have extensions in time. Kim's account can be refined to capture this fact by saying that an object x exemplifying a property p at duration of time t0-t1 causes an object y exemplifying property q at duration of time t2-t3. In short, event (x, p, t0-t1) causes event (y, q, t2-t3) instead. An alternative account from Richard Taylor states that an event can be defined as a change (or persistence) in the state (or states) of a substance (or substances) at a certain time (or duration of time). Both causes and effects are either changes in the states or the persistence of states of substances at particular times (or durations of time). The standard event causal account is

a direct relationship between changes of states or persistence of states, and only an indirect relationship between the substances themselves to which those states belong. One limitation of Taylor's account is prioritising of changes and persistence of states over the substances that bear them. Though it is far from true that substances determine their states in general, it is plausible to claim that substances provide the necessary material and structural conditions of bearing certain states. Keeping in mind the limitations of Kim's and Taylor's accounts of events, the event causal account I favour takes the event causal account as a direct relationship between the objects exemplifying certain properties (or substances in certain states) at certain times (or durations of time), as illustrated below:

| | | | | |
|---|---|---|---|---|
| …Event a | → | Event b | → | Event c… |
| Object x exemplifying | | Object y exemplifying | | Object z exemplifying |
| Property p at | | Property q at | | Property r at |
| Time t-2 to t-1 | | Time t0 to t+1 | | Time t+2 to t+3 |

It is generally accepted that only events can be the relatas of causation. According to the standard event causal account, objects have causal powers in virtue of their properties. When an object X possesses property P in some circumstance C at some earlier time (or duration of time), P causes (either deterministically or indeterministically) a certain effect at some later time (or duration of time). 'Causes' here can mean necessitate (if deterministic) or make probable (if indeterministic). Hence, P necessitates or makes it probable for X to cause Q in Y, given C. O'Connor, following Sydney Shoemaker, calls this 'simple functions from circumstances to effect' (Shoemaker 1984, pp.212). One limitation of Shoemaker's account is to take only properties, but not also objects as the bearers of causal powers. If we take objects as the necessary material and structural conditions of possessing certain properties,

then we can take the object together with the properties it possesses as one entity that has causal powers. Keeping in mind the limitation of Shoemaker's account, the event causal account I favour takes objects exemplifying properties as having causal powers. When an object X exemplifies property P in some circumstance C at some earlier time (or duration of time), XP causes (either deterministically or indeterministically) a certain effect at some later time (or duration of time). Hence, XP necessitates YQ or makes YQ probable, given C. In cases of an agent causing an action, the causes of actions and the actions themselves (effects) are events involving the agent. It consists of an agent (as substance) exemplifying certain properties (or states) at some earlier time (or duration of time) causing himself (as substance) to exemplify certain properties (or states) at some later time (or duration of time).

For O'Connor, proponents of agent causation are committed to the non-reductive (realist) causal relation between an agent and an event. Agents are substances that directly and purposively initiate, originate, produce, or bring about an event. It is therefore a direct relationship between the substances themselves and events, as illustrated below:

| | | | | |
|---|---|---|---|---|
| Agent x | → | Event a | → | Event b… |
| | | Object x exemplifying | | Object y exemplifying |
| | | Property p at | | Property q at |
| | | Time t0 to t+1 | | Time t+2 to t+3 |

This direct relationship between the substances themselves and events can be given further analysis. When agents exercise their unique causal powers (what O'Connor calls 'volitional-enabling active powers' or 'choice-enabling active powers'), they can be said to exemplify a certain property at a certain time that brings about events, as illustrated below:

| | | | | |
|---|---|---|---|---|
| Agent x | → | Event b | → | Event c… |
| Exercises Volitional Enabling | | Object y exemplifying | | Object z exemplifying |
| Active Powers at | | Property p at | | Property q at |
| Time t-2 to t-1 | | Time t0 to t+1 | | Time t+2 to t+3 |

Note that O'Connor does not make any claim concerning the temporal conditions of agent-causes. I am just assuming that agents as substance can exercise volitional-enabling active powers as and when he wants it, which I believe is consistent with O'Connor's view. When an agent X possesses property P (volitional-enabling active powers) at some earlier time (or duration of time), P neither necessitates nor make probable, but merely makes possible a certain effect at some later time (or duration of time). Hence, X's possession of P makes it possible for it to cause Y directly. It is worth emphasising that in contrast with Taylor's and Shoemaker's views, it is X (substance) alone, and not P (states or properties), that causes Y (action). In cases of an agent causing an action, the cause of action is solely the agent (as substance) whereas action itself (effect) is an event involving the agent. According to O'Connor, insofar as an effect is directly and purposively caused by an agent, agent causation cannot be described as 'simple functions from circumstances to effects' (O'Connor 1995, pp.177).

So, here lies the glaring difference between event causation and agent causation. On the event causal account, all events have antecedent causal conditions, whether deterministic or indeterministic. That is, there are antecedent causal conditions for objects to exemplify certain properties at earlier times (or durations of time), and certain properties exemplified by objects necessitate or make probable certain events at later times (or durations of time). On

the agent-causal account, there are no antecedent causal conditions for when and why agents exercise their volitional enabling causal powers, and the volitional enabling causal powers exemplified by agents merely make possible, but neither necessitate nor make probable events at later times (or durations of time). Since it is entirely up to the agents whether or not to exercise their unique causal powers, we can refer to the agents, and not their unique properties or causal powers, as the source of direct and purposive causes of events.

**Objections to Agents (Substance) as Causal Relata in Agent Causal Accounts**

O'Connor (1995) develops his agent causal account by replying to three objections to agents (Substance) as Causal Relata in agent-causal accounts, as advanced by Donald Davidson (1980), C.D. Broad (1952) and Carl Ginet (1990) as well as Roderick Chisholm (1971) and Richard Taylor (1974).

a. Davidson's Objection

Davidson poses a dilemma for agent causation in the following way: either causing by an agent of a primitive action is a further event, distinct from the primitive action, or it is not. The first horn of the dilemma can be expressed as: if the causing by an agent is a further event, then it is either an action or it is not. If it is a further action, then the causing by an agent is not primitive. If it is not a further action, then we have the absurdity of 'a causing' that is not 'a doing'. One way to avoid the first horn of the dilemma is to say that causing by an agent is not an event distinct from the primitive action. That is, 'causing by the agent' is constitutive of the primitive action. This brings us to the second horn of the dilemma, which can be expressed as follows: if causing by the agent is not a further event distinct from a primitive action, then the concept of cause plays no role and action is left unexplained.

O'Connor re-formulates the dilemma in terms of the relation between the agent as substance (a direct cause) and the event constituents of a primitive action. Now, the first horn of the dilemma, which says that 'causing by the agent' is a further event distinct from 'primitive action', can be resolved by saying that 'causing by the agent' is a direct cause by the agent as substance and not a further event. And the second horn of the dilemma, which says that the concept of cause plays no role and that action is left unexplained, can be resolved by saying that the agent as substance plays the causal role and his reasons play the explanatory role. The agent (which is not itself an action or an event) initiates, originates, produces, or brings about his action, under the guidance or influence of his reason (which is again not an action or event). So according to the agent causal account, agent is 'that which causes' intention-action and reason is 'that which explains' intention-action.

There are problems with O'Connor's reply. On the event causal account, objects (or agents) exemplifying certain properties (intentional states) at an earlier time (or duration of time) are the causes, and objects (or agents) exemplifying certain properties (action) at a later time (or duration of time) are the effects. Note that the definition of reasons as the content or object of the agent's intentional states is compatible with the event causal account. O'Connor wants to say that intention constitutes action because, first, he wants to show that no prior intentional states (as causes) are required to explain intention-action (as effects), and second, he wants to show that agent causation is more satisfactory compared to event causation in explaining action. This is because agents as causes and reasons as non-causal explanations need not appeal to antecedent conditions, and hence support absolute free will and ultimate responsibility. In contrast, events like agents exemplifying intentional states as causes and explanations always appeal to antecedent conditions, and hence do not support absolute free will and ultimate responsibility. However, when we appeal to agents as causes and non-

causal reasons as explanations of intention-action, both agents as causes and non-causal reasons as explanations are not themselves accounted for. The crux of the problem is this: when an agent acts for considered and deliberated reasons, does the agent as cause involve any exemplification of properties or change of states in the agent? If it does, then the agent's exemplification of properties or change of states at certain times (or durations of time) is the event that causally explains the agent's action. The agent's desires, beliefs, and intentions are 'properties exemplified by' or 'states possessed by' the agent. And reasons are the 'contents' or 'objects' of desires, beliefs, and intentions possessed by the agent. If it does not, then it seems that agents and their non-causal reasons have no causal and explanatory connections with our prior or existing psychological states that are causally and explanatorily relevant to our actions. For it can be asked what is it about the agent, if not his exemplification of properties or change of states at certain times (or durations of time), that causes his action? And it can be further asked how do reasons, if they are not connected to 'properties exemplified by' or 'states possessed by' the agent, explain his action? It seems that agents are uncaused causers and non-causal reasons are unexplained explanans.

A crucial element of O'Connor's reply is that non-causal reasons explanations supplement agent-causes when it comes to seeking a complete explanation of action. According to him, reasons are not reducible to events (agents exemplifying intentional states with reasons as their content at certain times or durations of time) and therefore are not causes. Since reasons are not causes, and since causal explanations are expressions of causal relations, they neither play causal roles nor figure in causal explanations. But it can be questioned whether reasons make a difference to, counterfactually supports, contribute towards, or are INUS conditions of action? If they are, then they are causes. And if they are not causes, or if they make no difference to, do not counterfactually support, do not contribute towards, or are not INUS

conditions of action, in what sense can we say that they guide or influence the agent's actions? Of course, O'Connor can reply that even though the agent is guided or influenced by his reasons, he need not act on those reasons. This means that despite the guidance or influence of the agent's considered and deliberated reasons, he can always act otherwise. If this is the case, then it seems that reasons do not explain the agent's action at all. This leaves the agent as substance as the sole explanation. Since the agent can always act otherwise despite his considered or deliberated reasons, it seems that his action is arbitrary or random, and hence unexplained or unmotivated. Note that even if causal indeterminism is true and the agent can always act otherwise, it remains true that the agent constrained by his considered or deliberated reasons and has no control over luck or random factors. To claim that an agent having a particular reason explains the agent's action is to claim that the agent acted because of that particular reason. But if a having a particular reason guided or influenced the agent, or if the agent considered and deliberated on a particular reason, but did not act on that particular reason, then the agent did not act because of having that particular reason, and having that particular reason does not explain his action. At most, non-causal reasons can be rationalisations and justifications of action, but not genuine explanations in the difference-making, counterfactual supporting, contributory, or INUS sense. That is, non-causal reasons are reasons that the agent may possibly act for (justifications and rationalisations), but not the reasons the agent actually acts on (explanations and motivations). The reasons that agents actually act on (explanations and motivations) are best conceived as difference-makers, counterfactual supporters, contributors, INUS conditions, or as causes in short, of his action. So, even if we grant that O'Connor succeeded in responding to Davidson's dilemma, he does so at the price of stipulating the agent as an uncaused causer who can act otherwise despite the guidance and influence of his considered and deliberated reasons, which leaves his action arbitrary and random, and hence unmotivated and unexplained.

b. Broad's & Ginet's Objections

An objection, initially raised by Broad and later developed by Ginet, can be expressed as follows: When I caused an intention at time t, I can neither explain why I decided when I did, nor why I decided as I did. Event causation can easily account for this: exemplifications of properties (intentional states) by objects (agents) at an earlier time (or duration of time) cause and causally explain the exemplifications of properties (action) by objects (agents) at a later time (or duration of time). Agent causation holds that agents are substances or continuants that are said to have direct and purposive causal powers with no appeals to further event-causes at earlier times to explain the exemplification of properties by substances or continuants (or objects) at a later time.

The agent-causal account accepts that agent-causes alone are insufficient for explaining the agent's acts, and they must be supplemented by the agent's reasons. In other words, agent-causes can partially, but not completely, explain when and why an agent decided to do as he did. The agent's reasons complete the explanation. However, the problems associated with using reasons, conceived non-causally, to fill the explanatory gap remain. Given that reasons are not conceived as causes under O'Connor's agent-causal account, they cannot figure in causal explanations. Consequently, the causal explanations of when and why an agent decided to do as he did have nothing to do with reasons, conceived non-causally. Agent-causalists may reply that reasons need not be conceived causally to complete the explanation. But it can be further questioned whether explanans that make no differences to, do not counterfactual support, do not contribute towards, or are not INUS conditions, can explain when and why an agent decided to do as he did. Again, the distinction between acting for possible reasons and acting on actual reasons is crucial here. Non-causal reasons are the possible reasons an agent acts for, and their role is the justification and rationalisation of

action. Causal reasons are the actual reasons an agent may act on, and their role is the explanation and motivation of action. Non-causal reasons become causal reasons when one or more of the possible reasons the agent may act for (justifications and rationalisations) becomes the actual reason or reasons the agent acts on (explanations and motivations). Only causal reasons; reasons the agent actually acts on; reasons that make differences, counterfactually support, contribute towards, or are INUS conditions of action; can explain when and why the agent decided to do as he did. So, O'Connor's reply to the Broad-Ginet objection only works to the extent that it adds non-causal rationalisations and justifications to the agent-causal ones, and it does not complete the explanation. The agent causalists now faces a dilemma: either insist that reasons do not figure as causes and leave when and why an agent decided to do as he did unexplained or unmotivated, or concede that reasons do figure as causes and leave the causal role of the agent as substance reduced or eliminated.

c. Chisholm's & Taylor's Objections

In their earlier writings, both Chisholm and Taylor were agent causalists. In their later writings however, they came to believe that that agent causation is a subspecies of event causation, and raised objections against the agent causal account. For Chisholm, agent causation is a subspecies of event causation, and defines sufficient causal condition as follows: S is a sufficient causal condition of E = S is a set of properties such that it is a law of nature that, if all members of S are exemplified by the same object at the same time, then E will be exemplified either at that time or later. But Chisholm objects that the agent causal account presupposes that when an agent contributes partially to the cause of E, the agent-cause itself has no further causes. If this is the case, then it seems that the occurrence of the agent-cause is completely arbitrary and random. And it seems that the agent as substance cannot be held morally responsible for the agent-cause to occur. The problem with agent-

causal accounts is that agents are substances that can be direct causes of events, without themselves having antecedent sufficient conditions. Taylor attempts to correct this by postulating that agent-causes can have sufficient causal conditions for their occurrences. All that agent causal accounts require is that agents are the initiators, originators, or producers of agent-causes, and it does not matter whether agent-causes are themselves causally determined or not.

O'Connor replies that the idea of there being antecedent sufficient conditions for agent-causes is unintelligible. In event causation, the cause of 'event A causing event B' is none other than the cause of event A. In agent causation, it makes no sense to say that the cause of 'agent A causing event B' is the cause of agent A. This is because there is a substance, and not event, at the front end of the agent causal structure. Since the agent causal structure terminates with a substance at the front end without preceding events it is impossible for there to be antecedent sufficient conditions for an agent-cause. So, Chisholm's criticism that the agent as substance cannot be held morally responsible for agent-cause because they can have no antecedent sufficient conditions, and Taylor's criticism that agent-causes can have antecedent sufficient conditions are clearly mistaken.

Before evaluating O'Connor's reply, let's look at his agreements and disagreements with Chisholm and Taylor first. O'Connor agrees with Chisholm that agent-causes can contribute partially to the occurrence of an action (and perhaps adds that agent-causes should be completed by the agent's reasons), and that agent-causes may be a necessary but not a sufficient condition for the occurrence of an action (and perhaps adds that agent-cause requires agent's reasons to be sufficient). He disagrees with Chisholm that the agent-cause itself must have a sufficient causal condition. O'Connor agrees with Taylor that there can be antecedent sufficient conditions for the occurrence of an action, and adds that agent-causes

(together with reasons explanation) can be the antecedent sufficient conditions for the occurrence of an action. In short, they both agree that action can be determined, and O'Connor adds that the agent as substance (together with his reasons) can determine his action. He disagrees with Taylor that the agent-causes themselves have antecedent sufficient conditions, or that agent-cause is itself causally determined. It is one thing to say that agent-causes do not have antecedent sufficient causal conditions, quite another to say that agent-causes do not have antecedent causal conditions at all. O'Connor's position seems to affirm the latter, and by implication, the former as well.

So, O'Connor's main disagreement with both Chisholm and Taylor is that there cannot possibly be any antecedent sufficient causal conditions for agent-causes. The reason that he cited is the internal causal structure of agent causation, where an agent (a substance instead of an event) is posited at the antecedent end, which makes it a 'doing' rather than a 'causing'. 'Doing' by definition is just what the agent initiates, originates, produces, or brings about. 'Causing' is just another occurrence (an event) with further causes. Agent causation not only precludes the possibility of it having antecedent sufficient causal conditions, but also precludes the possibility of it having any antecedent causal conditions. Put simply, agent-causes are themselves uncaused.

The problem with O'Connor's reply seems to rest on a problematic presumption that the agent as substance is the direct initiating and originating cause of his action without the need to appeal to his prior or existing intentional states. This implies the agent as substance bears no causal relation with his prior and existing intentional states at all. O'Connor admits that there is no embarrassment in acknowledging that the agent-causes are themselves uncaused. For Chisholm, uncaused agent-causes are arbitrary and random, and hence are inexplicable or unintelligible. Arbitrary and random, and hence inexplicable or unintelligible agent-causes

cannot be used to ground moral responsibility. To answer the arbitrariness and randomness charge, O'Connor's account supplements agent-causes with non-causal reasons explanations, such that agent-causes are made explicable or intelligible by having reasons guide or influence the agent. But O'Connor further admits that the agent's reasons are not causes and they are not causally related to the agent and his action. Since the agent's reasons are not causally related to the agent and his action, they make no differences to, do not counterfactually support, do not contribute towards, or are not INUS conditions of, the agent and his action. It is unclear how the terms 'guidance' or 'influence' can be understood apart from the causal sense. Perhaps the agent's reasons guide or influence him by being factors taken into consideration during his deliberation. But the question whether they explain the agent's action remains. For it seems that the reasons considered during the agent's deliberation can hardly explain his action unless they make a difference to, counterfactually supports, contribute towards, or are INUS conditions of his action. That is, stating the possible reasons the agent acts for merely justify or rationalise but does not explain his action; only stating the actual reasons the agent acts on does so. If the actual reasons the agent acts on make no differences to, do not counterfactually support, contribute toward, or are INUS conditions of his action, then they do not explain his action. And if the actual reasons the agent acts on do not explain his action, then it seems that agent-causes are simply stipulations which may be completely arbitrary or random and are hence vulnerable to the charge of inexplicability or unintelligibility. So, it seems that Chisholm's worry that the uncaused agent as substance cannot be used to ground moral responsibility remains.

**Objections to Non-Causal Reasons Explanations in Agent-Causal Accounts**

As discussed above, O'Connor believes that the causal theory of reasons explanation, where reasons directly influence actions by causally producing them, is mistaken. He argues for a non-causal theory of reasons explanations instead, which he believes is neither incompatible with nor an alternative to agent causation. In the non-causal theory of reasons explanations, agent causation is necessary to play the causal role while reasons are necessary to play the explanatory role in a complete causal explanation of action. So, the agent's actions can be explained by citing and ascribing reasons for which the actions were performed. O'Connor (1995) further develops his agent causal account by replying to three objections to non-causal reasons explanations in agent-causal accounts, as advanced by Galen Strawson (1986), Robert Kane (1989), and Stewart Goetz (1988).

a. Strawson's Objection

Strawson objects that the non-causal conception of an agent acting from prior reasons and yet not causally determined by them leads to a vicious regress. An agent deciding whether to act in accordance with reason X or reason Y only if there are some further reasons that decisively incline him either way.

O'Connor replies that reason explanations need not appeal to second-order (or higher-order) reasons to make first-order reasons intelligible. Given actions X and Y, and corresponding reasons X and Y, citing reason X would be sufficient to explain why an agent chose action X, even when reason X does not causally produce action X, and he could have chosen differently under exactly the same circumstances.

There are problems with O'Connor's reply. First, when O'Connor says that citing reason X would be sufficient to explain why an agent chose action X, does he mean sufficient causal

condition, sufficient causal explanation, or sufficient non-causal explanation? What is needed here is a sufficient causal condition or sufficient causal explanation of X. Second, O'Connor says that reason X does not causally produce action X. So, reasons cannot be a sufficient causal condition. And since a causal explanation expresses a causal relation, a reason cannot be a sufficient causal explanation as well. The remaining option is reasons as sufficient non-causal explanations, which does not provide the causal condition or causal explanation of action. Third, O'Connor says that even if reason X is sufficient to explain action X, the agent could have chosen action Y under the same circumstances. If this means that it is possible for the agent to choose action Y despite citing reason X, then reason X cannot be a sufficient explanation, non-causal or otherwise. If this is the case, then it seems that Strawson's point about the need for reasons explanations to cite some further reasons that decisively incline him either way still stands.

b. Kane's Objection

Kane objects by asking what makes an agent choose action X, given exactly the same past and the same reasons, rather than the alternatives (like action Y). His answer is that citing all the salient psychological features of the agent at the time of his intention is necessary and citing reason X alone is insufficient. O'Connor replies that presence of both non-causal reasons and agent-causes are necessary and jointly sufficient components in the agent causalists' explanatory scheme.

Both Kane and O'Connor agree that reasons explanations are not sufficient to explain action. Something more is required. Their disagreement lies in the ways they fill the causal gap. Kane appeals to the salient psychological features of the agent at the time of his intention. This is consistent with event causation. An object (the agent) exemplifying properties (intentional states) at an earlier time (or duration of time) causes or explains the object (the

agent) exemplifying properties (action) at a later time (or duration of time). Of course, there are prior event-causes and explanations at some earlier time (or duration of time) as to why the agent exemplifies the salient psychological features at some later time (or duration of time). Kane can accept that these salient psychological features have their own causes only if they are inherently indeterministic. O'Connor appeals to direct causation by the agent as a substance. This is inconsistent with event causation. The agent causes his action directly without being caused by any earlier events (including the agent's own intentional states). As discussed, agent-causes are themselves uncaused, and this means that they are causally inexplicable stipulations, which may be completely arbitrary and random. Both Kane's and O'Connor's solutions to fill the causal gap face the problem of arbitrariness and randomness in different ways. Kane's solution faces the luck problem because the event-causes of actions are causally indeterministic whereas O'Connor's solution faces the luck problem because the agent-causes of action are themselves uncaused. The choice here seems to be between a causally indeterministic event causal solution and an uncaused agent-causal solution. It seems that O'Connor's account fares no better than Kane's account in facing the problem of arbitrariness or randomness.

c. Goetz's Objection

Goetz objects that the agent causing of his action can only be explained by appealing to the reason for which he acts. Any explanatory power which the agent's causing of his action may have is 'derived from' or 'parasitic upon' the explanatory power of the reason for which he acts. Thus, the agent's causing of his action is redundant for this explanatory role once the reason for which he acts is cited. For O'Connor, the relative importance of the explanatory powers between the agent's causing of his action and the reason for which he acts is the converse of Goetz's suggestion. The agent's free exercise of his causal capacity provides a

necessary link between reason and action, without which reasons could not explain the action in any significant way. It allows reasons to guide or influence action without causing it. If the agent-cause is removed, the causal link between reason and action disappears.

One alternative to agent causation in explaining the causal link between reason and action is the causally indeterministic theory of reasons explanation, where all the salient psychological features including reasons directly cause actions without determining them (Kane's account). O'Connor does not deny that this alternative is a viable one, but rejects it on grounds that it fails to show how it can be entirely up to the agent to determine which among a range of possible courses of action he will actually undertake. It may be true that Goetz's attempt to collapse agent causal explanation into reasons explanation does not succeed because he does not provide any causal account of action. Agent causation cannot be reduced to reasons explanation precisely because agent causation provides a causal account of action and makes reasons explanation of action possible and even significant.

Both Goetz and O'Connor agree that the agent's reasons are what make his action explicable and intelligible. I agree with O'Connor but not Goetz that since the agent's reasons are not causes, a causal account of action is required and that the agent is indispensible in any causal account of action. I can even agree with O'Connor (but not Kane) that the agent causal account (but not the event causal account) succeeds in showing how it can be entirely up to the agent to determine which among a range of possible courses of action he will actually perform. My disagreement with O'Connor is in the treatment of the agent and the reasons he actually acts on. First, O'Connor agrees that the agent as substance on its own cannot make his action explicable or intelligible and that reasons are required to do so. Second, I believe that the reasons that make the agent's action explicable and intelligible are the reasons he actually acts on and that such reasons have to make a difference to, counterfactually support,

contribute towards, or are INUS conditions of the his action. In short, the agent's reasons have to figure as causes of action. O'Connor may agree with the former but not the latter. Third, if reasons that make an agent's action explicable or intelligible are only those that make a difference to, counterfactually support, contribute towards, or are INUS conditions of the his action, then those that do not fail to make the agent's action explicable or intelligible. If my argument is acceptable, then the conjunction of agent-cause and non-causal reasons explanations does not make the agent's action more explicable or intelligible. Even if it is true that the agent-causal account (but not the event-causal account) succeeds in showing how it can be entirely up to the agent to determine which among a range of possible courses of action he will actually perform, the agent-cause itself remains an inexplicable or unintelligible stipulation. And even if O'Connor correctly insists against Goetz that actions must have causes, what he postulates as the cause of actions remains an inexplicable or unintelligible stipulation.

**Objection to Agent Causation as Mere Randomness**

The objection to agent causation as mere randomness can be expressed as follows: it seems that it is in principle impossible for us to ever to know whether any events are caused in the way the agent causal account postulates, because such an event would be indistinguishable from one which is essentially random, not connected to events preceding it. Put another way, it seems that we can never know whether the property that gives rise to agent-causes is instantiated. This is so because unlike event causation, where objects instantiating certain properties at some later time (or duration of time) are connected to objects instantiating certain properties at some earlier time (or duration of time), agent causation postulates that objects (agents) initiate, originate, produce, or bring about subsequent events without any

antecedent conditions. Agent causalists may claim that agents do instantiate volitional enabling active powers at certain times (or durations of time) to cause an event. But this merely pushes the objection of mere randomness one step back. For according to O'Connor, the volitional enabling active powers instantiated by the agent neither necessitates nor makes it probable that the agent causes the event; and the exercising of volitional enabling active powers is entirely up to the agent. So, the agent may causally determine an event without himself having any antecedent causal conditions, and is hence an uncaused causer.

Again, O'Connor replies by appealing to the role of reasons in agent-causal accounts. Reasons are not causes and so require a causal link that relates the agent's reasons to his actions. The agent's capacity to initiate, originate, produce, or bring about an action, guided and influenced by his consciously held reasons, is an appropriate candidate for this causal link. Without the mediation of the causal link, the agent's reasons alone cannot explain his intention and action. The agent causal account holds that both agent-causes and reasons explanations are jointly sufficient conditions for action.

There are problems with this reply. As seen in his reply to Goetz above, O'Connor argues that reasons are necessary but not sufficient for explanations of the agent's action, and agent-causes are required to provide the causal account that is lacking in non-causal reasons explanations. But as seen in his reply to Broad and Ginet above, O'Connor argues that agent-causes are necessary but not sufficient for explanations of the agent's action, and non-causal reasons are required to explain when and why the agent decided to do as he did, which is lacking in agent-causes. And he wants to show that both agent-causes and reasons explanations are jointly sufficient conditions for action to escape the charge of mere randomness. The problem is that when we observe the roles played by agent-causes and reasons explanations in the agent-causal account, an explanatory gap remains. While the

agents as substance play the causal role, non-causal reasons explanations play only the justificatory or rationalising role. Justifications or rationalisations and explanations can come apart. In other words, what justifies or rationalises may not be what explains. O'Connor holds that the agent can act otherwise despite the guidance or influence of his considered and deliberated reasons. If the agent can act otherwise despite the guidance or influence of his considered and deliberated reasons, then these reasons are those that the agent can possibly act for, but that are not acted on, do not explain his actions, even if they justify or rationalise them. Hence, non-causal reasons explanations do not close the explanatory gap and agent-causes remain inexplicable and unintelligible.

And as seen in the discussion of O'Connor's reply to Davidson, the problem can be expressed as follows: when an agent acts for considered and deliberated reasons, does the agent-cause involve the agent's exemplification of properties or change of states (having certain desires and beliefs) with certain content (reasons) at certain times (or durations of time)? If it does, then the agent's exemplification of properties or change of states (having certain desires and beliefs) with certain content (reasons) at certain time (or durations of time) are events that necessitate or make probable the agent's intention and action. And if this is true, then the direct and purposive agent-causes can be causally explained by the agent's exemplification of properties or change of states (having certain desires and beliefs) with certain content (reasons) at certain time (or durations of time) and this makes direct agent causation redundant. If it does not, then it seems that agents and reasons are respectively uncaused causes and unexplained explanans that have no causal and explanatory connections with our prior or existing psychological states. Again, it can be asked what it is about the agent, if not the exemplification of properties or change of states, that causes his action. And it can be further asked how do reasons, if not conceived as part of the 'properties exemplified by' or

'states possessed by' the agent, explain his action? On the agent-causal account, agents (as substances) have no causal connections with our prior or existing psychological states and hence are uncaused causes, and reasons have no explanatory connections with our prior or existing psychological states and hence are unexplained explanans. Uncaused causers or unexplained explanans are vulnerable to the mere randomness objection. Hence, the agent-causal account does not succeed in avoiding the mere randomness objection.

**Reasons as Elevators of Objective Propensities and Tendency-Conferring States**

The discussion so far focuses on the earliest of O'Connor's work – the paper 'Agent Causation' (O'Connor 1995). His position seems to have changed a little in his later work – the book *Persons and Causes* (O'Connor 2000). In the more recent work, there is an inclination to accept some form of causal indeterminism, thereby admitting that agent-causes of action can have antecedent conditions or prior causes and that reasons are causes after all. First, he suggests that an agent's recognition of a reason to act induces or elevates an objective propensity to initiate the behaviour. In other words, agent causation is probabilistically structured not only by tendency-conferring states of having reasons to act in specific ways but also by more enduring states of character, involving relatively fixed dispositions and long-standing general intentions and purposes around which the agent's life has come to be organised. Second, he maintains that embracing this causal-propensity account of the relative strength of reasons – and perhaps even supposing it to be required to make sense of the very idea that reasons in general motivate actions – need not lead one to abandon the non-causal relationship between actions and the reasons that explain them. For the mere fact that a reason the agent had gave him some tendency in this sense to act as he did does not explain his action. Maybe he did not act for that reason, despite his recognition

that it was a relevant reason. And third, he holds that the agent is the sole causal factor directly generating his intention to act (not a co-cause along with the agent's reasons), but the agent doing what he does is shaped, causally, by his total motivational state.

I find some tensions in O'Connor's revised view. He accepts that the agent's intention and action are at least indeterministically caused or partially conditioned (made probable rather than merely made possible) by our momentary reason states and long-standing character, and that reasons and character constitute the total motivational state for an agent's action. And yet he denies that the relationship between reasons (as well as character) and action is causal, insisting that the agent as substance is the sole cause of his action. His argument for this conclusion is that reasons and character confer a tendency for an agent to act, but it does not necessarily explain the agent's action. For it is possible for the agent to recognise relevant reasons to act without acting on them. Perhaps O'Connor wants to say that reasons and character are indeed causally indeterministic antecedent conditions, but the agent can still exercise his unique causal powers to act or not to act despite those antecedent conditions. That is, causal indeterminism may be true of reasons and character on the agent, but the agent nevertheless possesses the powers of self-determination. But this account raises some problems. First, if it is accepted that reasons and character are probabilistic or statistical causes (tendency-conferring states that induce or elevate objective propensity) of action, then they are causally relevant to (make a difference to, counterfactually support, contribute towards, are INUS conditions of) an agent's action. So, why deny that they stand in some causal relationship with the agent's intention and action? Second, if it is insisted that the relationship between reasons (as well as character) and action is a non-causal one, then why admit that reasons (as well as character) are causally relevant to action? And why bother to treat reasons as tendency-conferring states that induce or elevate objective propensity? Third,

it seems inconsistent to insist that the agent (as substance) is the sole cause of his action and that the agent doing what he does is causally shaped by his total motivational state (reasons and character) at the same time. If the agent (as substance) is the sole cause of action, then the total motivational state is causally irrelevant to action. And if the total motivational state is causally relevant to action, them the agent (as substance) is not the sole cause of action.

**Agents and Reasons as Co-Causes of Action**

In his paper 'Agent Causation and Event Causation in the Production of Free Action', Randolph Clarke proposes an alternative agent-causal account whereby the agent-causes (as substance) and event-causes are co-causes of action (Clarke 2009). The alternative view agrees with the traditional view (like O'Connor's) that causation by agents is not reducible to, identical with, or consisting in causation by events. In particular, the alternative view agrees with the traditional view that while the agent causes an event, he himself is not an effect of prior event-causes over which he has no control over. However, the alternative view also agrees with critics of the traditional accounts that agent causation is mysterious at worst or unintelligible at best. Whereas the traditional view maintains that action has no prior event-causes and that the agent is the sole determinant of action, the alternative view maintains that action has prior but nondeterministic event-causes, and that the agent is the co-determinant of action. The alternative view holds that the rejection of event-causes of action is unnecessary and that the acceptance of nondeterministic event-causes of action helps to remove the mystery or unintelligibility problem faced by the traditional view. Clarke illustrates with the following example. Suppose there is a nonzero probability that a prior event R1 consisting in an agent's having (or acquiring) certain desires and beliefs will nondeterministically cause event A1, a certain action performed by the agent. And suppose there is also a nonzero

probability of a prior event R2 consisting in an agent's having (or acquiring) certain alternative desires and beliefs will cause event A2, a certain alternative action. Suppose further that whichever of the actions the agent performs, that action will be caused by the reasons that favour it, only if the agent causes that action. Finally, suppose the agent performs action A1, which is caused by him as well as by R1 nondeterministcally. On the alternative view, the agent's acting with free will consists in his action's being caused this way by him and by his reasons. Moreover, agent causation does not interfere with the event-causal explanation of action, for agent causation does not interrupt or divert the causal route from reasons to action. The explanation that cites the agent's reasons is in no way false or incomplete, even if it does not tell the entire story about what caused the action, and even if it does not reveal direct agential control as a necessary condition.

Clarke discusses the different varieties of control over action to explain the insufficiency of event-causes and the necessity of agent-causes for acting with free will. First, an agent is said to have bare actional control over his action when his action is caused in the right way by her reasons. Second, an agent is said to have rational control over his action when he has the capacity to reflect rationally on the alternatives available, the reasons favouring each, and to govern his behaviour on the basis such reflection. Clarke notes that an agent possesses both bare actional control and rational control even if all causes of his action are events, and even if all causes of his action are deterministic. However, both these varieties of control are not sufficient for free will as an agent may have the capacity for rational self-determination and yet fail to determine how the capacity is exercised or which alternative is selected. Hence, bare actional control and rational control are not sufficient for free will. And yet it appears that no further control is available to the agent if his action is nondeterministically caused by events and not by anything else, for nondeterministic event-causes merely leave room for but

do not by themselves constitute further control over which alternative is selected. Causation by the agent as a substance appears to make such further control possible. In agent-causing a certain action, the agent exercises a causal power distinct from the causal power of any events involving the agent, over which alternative is selected. According to Clarke, this third variety of direct agential control is necessary for free will, and sufficient when conjoined with bare actional control and rational control.

Clarke then explicates a widely held objection to his alternative account of agent causation that it fails to fulfil the following requirement for acting with free will: if any event causes a certain action, an agent acts with free will only if he controls the occurrence of that event. This implies that in acting with free will, an agent has to be the sole determinant of his action (like O'Connor's traditional account). But agents in general do not control their having (or acquisition) of reasons for action, and so the alternative view fails to fulfil this requirement. Clarke responds that this requirement is mistaken. Nondeterministic event-causes of action make it possible for an agent to act with free will, even if his action is caused by a certain event over which he does not have any control, as long as the agent exercises a causal power that influences whether that event is followed by a certain action. Clarke notes that the objection can be pushed further as follows: on the alternative view, the agent is not the only entity that influences whether a certain event will be followed by a certain action, for other prior events exercise such influence as well. And in order to act with the control that constitutes free will, the agent must be the sole entity exercising influence over whether these prior events will be followed by a certain action. Clarke responds that no argument has been given to support this further objection. And it appears evident that prior events influence what an agent does. But this fact alone is not sufficient to show that the agent lacks free will. Free will seems to allow for the influence of entities other than the agent on what the agent

the agent does, and on whether certain prior events are followed by an action. Clarke notes that traditional views attempt to capture this fact. On O'Connor's later view for example, prior events that influence what the agent does establish probability structures leaving some actions more likely than others, even if those prior events do not cause what the agent does. Clarke contends that there is no adequate non-causal construal of such influence. And even if there is, the agent is still not the sole entity that influences what he does. If entities other than the agent exercise genuine influence on what the agent does, it does not matter whether such influence is construed causally or not. If this is the case, then it seems that on the issue of control, there is no real difference between genuine non-causal influence and causal influence on what the agent does. If causal influence by prior events is inconsistent with free will, so does genuine non-causal influence. Conversely, if genuine non-causal influence by prior events is consistent with free will, so does causal influence.

In his 'Libertarian Views: Dualist and Agent-Causal Theories', O'Connor discusses Clarke's alternative view (O'Connor 2002, pp.353). He observes that Clarke's alternative view does not achieve the desired integration of his later view as it claims that reason has a tendency to produce action, and this seems to imply that action has some chance of occurring apart from the agent's activity (direct agent-causes). The alternative view seems to conceive reasons as actively competing with the agent to produce action. However, O'Connor also notes that on Clarke's view, an action will be caused by the reasons that favour it only if the agent causes that action, and this seems to imply the hegemony of agent-cause over the probabilistically structured tendencies conferred by reasons. O'Connor suggests that better integration can be achieved by treating reasons as genuine non-causal influences and agents as the sole cause of action. This avoids the competition between agent-causes and reason-causes. I agree with O'Connor's assessment but not his suggestion. Clarke claims that (1) an agent has no causal

influence or control over his reasons of action, or an agent does not interfere with, interrupt, or divert the causal route from reasons to action, (2) an agent has sole causal influence or control over whether an action follows from the reasons that favour it, and an action will be caused by the reasons that favour it only if an agent causes that action.

One problem with Clarke's alternative view is that claims (1) and (2) seem to be inconsistent (or at least in tension) with one another. If claim (2) is read as agent-causes solely determine which alternative action follows from the reasons that favour them, and hence have some influence or control over which of the reasons for action are effective, then this seems to be inconsistent (or at least in tension) with claim (1). Using Clarke's example above, it can be asked whether the agent can determine whether A1 follows from R1 or A2 follows from R2 despite their given probability structure. Or alternatively, it can be asked whether the agent is constrained by the probability structures of R1 to A1 and R2 to A2. If the agent can determine whether A1 follows from R1 or A2 follows from R2 despite their given probability structure, or if the agent is not constrained by the probability structures of R1 to A1 and R2 to A2, then R1 and R2 are hardly causes of A1 and A2 respectively, only the agent is. Here, it seems that the agent does interfere with, interrupt, or divert the causal route from reasons to action, or that the agent has causal influence or control over reason-causes of action. That is, claim (2) seems to be inconsistent (or at least in tension) with claim (1). One way to reduce the tension is to give up claim (1). This grants the agent sufficient control to act with free will but threatens agent having (or acquiring) reasons as co-causes of action. Here, O'Connor would be right to say that Clarke's view seems to imply the hegemony of agent-cause over the probabilistically structured tendencies conferred by reasons. Another way is to weaken claim (2) by saying that there is a nonzero probability that both the agent and his having (or acquiring) reasons co-cause an action, and that both the agent and his having (or acquiring)

reasons are necessary and jointly sufficient causal conditions of action. This grants the agent's having (or acquiring) reasons as genuine co-causes of action but the sufficiency of the agent's control to act with free will is threatened. Here, neither the agent-cause nor the probabilistically structured tendencies conferred by reasons has hegemony over one another. Hence, Clarke's view faces the following dilemma: either affirm that an agent has sufficient control to act with free will and reject that an agent having (or acquiring) reasons are genuine co-causes, or affirm that an agent having (or acquiring) reasons are genuine co-causers and reject that an agent has sufficient control to act with free will.

Another problem with Clarke's alternative view is identified by O'Connor as the integration problem but I will approach it differently and draw a different conclusion. Other than claims (1) and (2) , Clarke further claims that (3) an agent is <u>not</u> an effect of prior event-causes over which he has no control, or an agent-cause has <u>no</u> antecedent causal conditions. This claim is a necessary premise for any agent-causal account. To weaken or give up claim (3) is to abandon any form of the agent causal account. Claim (1) suggests that 'an agent's direct causing of his action' has <u>no</u> causal influence or control over 'his having (or acquiring) prior reasons for action'. Claim (3) suggests that 'an agent's having (or acquiring) prior reasons for action' has <u>no</u> causal influence or control over 'his direct causing of his action'. Both claims seem to strongly suggest that an agent (as substance) and his having (or acquiring) reasons are separate, distinct, independent, and unrelated causes of action. It is highly counterintuitive to treat the agent (as substance) and his having (or acquiring) reasons, which seems so intimately related, as separate, distinct, independent, and unrelated causes of action. Even if they are not separate, distinct, independent, and unrelated, it is still difficult to see how the agent and his having (or acquiring) reasons can be causally integrated. A further problem lies in the mystery and unintelligibility of claim (3) given that an agent and his

having (or acquiring) reasons are separate, distinct, independent, and unrelated. It can be asked what explains why the agent acts on one reason rather than another. It cannot be further reasons for it contradicts claim (3). It cannot be character too for it contradicts claim (3) as well, which states that an agent is not an effect of prior event-causes over which he has no control, or an agent-cause has no antecedent causal conditions. And if total motivational state is constituted by reasons and character, then it follows that the total motivational state cannot be the answer as well. The remaining option seems to be that nothing explains why the agent acts on one reason rather than another. If this is the case, then the mystery or unintelligibility problem that afflicts O'Connor's traditional view affects Clarke's alternative view as well.

**Event Causation versus Substance Causation**

The objections I raised against O'Connor's and Clarke's accounts so far focus on the relation between the agent and reasons components of the agent causal explanation of action. On O'Connor's earlier account, it is hard to see how the agent's reasons explain his action if it does not play a causal role (at least in the difference-making, counterfactual-supporting, contributory, or INUS sense). On O'Connor's later account, it seems implausible to affirm reasons as tendency-conferring states that induce or elevate objective propensity and yet deny them a causal role (at least in the probabilistic and statistical sense) at the same time. On Clarke's account, it is hard to see how reasons can be indeterministic co-causes of action when they play no causal role in co-determining which available alternative action is taken. Both O'Connor's and Clarke's accounts eventually fall back on the agent as substance as the ultimate explanation of action. So far, I have only mentioned that agent-cause (whether as sole or co-cause), and not event-cause, as the ultimate explanation of action is inexplicable or

unintelligible, but I have never argued why it is implausible, if not impossible, for agent as substance to figure as a sole or co-cause. I will be attempting this next.

Responding to Broad's objection in his later paper 'Agent Causal Power', O'Connor questions the claim that events are ontologically prior to substances, and that substances existing at a time depends metaphysically upon events (that involve them) existing at a time. He suggests a more symmetrical alternative where substances and events are not ontological prior to one another, or where substances and events are metaphysically dependent upon one another. He expresses it as: necessarily, an object O at time t exists only if there is some event E that involves it occurs at time t, and for any event E that occurs at time t that involves object O, necessarily, event E occurs only if object O exists at time t. (O'Connor 2009, pp.377). For Clarke, causal relatas are concrete particulars, and both substances and events are particulars. If events can be causes but not substances, there must be some reason for this difference. It is not adequate to claim that the implausibility, if not impossibility, of substances as causes is self-evident. And as there are no adequate accounts of this difference so far, the possibility of causation by a substance remains unresolved (Clarke 2009, pp.347).

In chapter six of his book *Personal Agency: The Metaphysics of Mind and Action* (Lowe 2008), E.J. Lowe carries O'Connor's and Clarke's arguments further. He claims that agent causation is irreducible to event causation, and even suggests that event causation is reducible to agent causation. Lowe takes agent to mean a persisting object or an individual substance in the broad sense (whether animate or inanimate) that possesses various properties, including certain causal powers and liabilities. A causal power is a power to cause some object to act in a certain way, for example, a power to cause sugar to dissolve. And a causal liability is a power to be caused by some object to act in a certain way, for example, a power to be caused to dissolve in water.

In arguing for the conceptual priority of agent causation, Lowe observes that only a creature capable of intentional action can acquire knowledge of causal relations between events from experience. Such a creature must not only be an agent, but must also be *aware* of being an agent in order to possess the concept of agent causation. Because we are causal agents and are aware of our ability to intervene in and manipulate the course of nature, we are able to test causal hypotheses experimentally and thereby distinguish (not infallibly) between causal and non-causal events. An important component of this ability to intervene in the course of nature is our capacity to move *ourselves* at will in accordance with our desires (within limits) rather than being moved entirely by external factors. But a purely passive creature, however acute its powers of observation, would be incapable of discriminating empirically between causal and non-causal sequences of event, and hence it seems reasonable to conclude that it would have no concept of event causation. The concept of agent causation cannot be derived from that of event causation because possession of the former concept is a prerequisite of possession of the latter concept. It cannot be that we *first* learn to apply the concept of event causation to observable events and *only then* learn to conceive of ourselves as agents.

In arguing for the ontological priority of agent causation, Lowe claims that *events of themselves possess no causal powers and liabilities*. Only *persisting objects* or individual substances do. It is such entities that we describe as being magnetic, corrosive, inflammable, soluble, and so forth. Objects and substances manifest or display their causal powers and liabilities by acting on things, or being acted upon, in various appropriate ways, by attracting, corroding, burning, dissolving, and so forth. Agent causation should be conceived of as a species of substance causation. In describing such activities we use the language of agent causation rather than the language of event causation. We resort to the latter when we are at least partially ignorant about the causal agents that are at work. For example, what is it for

water to dissolve salt? It is, precisely, for water to *cause* salt to dissolve. For *x* to dissolve *y* is for *x* to cause *y* to dissolve, just as for *x* to move *y* is for *x* to cause *y* to move. Dissolving and moving are species of *causation* and the entities that engage in these species of causation are individual *substances* or persisting objects of various kinds. It is plausible to say that whenever a substance causes an event, it does so by acting in a certain manner and that its acting in such a manner constitutes an event. So, rather than the reducibility of substance causation to event causation, it seems that the very reverse is true.

According to Lowe, the problem of free will is intractable because the contemporary free will debate is conducted in terms of a mistaken approach to causality in general, which assumes that all causation is fundamentally *event* causation. An event must either *have or lack a further event cause* and we are faced with a dilemma. For if our choices are *caused*, we lack genuine freedom and responsibility. And if our choices are *un*caused, they seem to be mere chance happenings, and we lack genuine freedom and responsibility too. Agent causalists can resolve this dilemma by claiming that agent-cause is itself uncaused without implying that agent-cause is a mere chance happening: for agent-cause is not a *happening*, an event, and is therefore not the sort of thing to have an event as a cause.

I want to raise four objections against Lowe's arguments for the irreducibility and priority of substance causation. First, there seems to be a contradiction between Lowe's claim that a persisting object or an individual substance possesses certain causal liabilities, which is a power to be caused by some other object or substance to act in a certain way, and the claim that an agent-cause (which Lowe admits as a species of substance-cause) is itself uncaused. Perhaps he wants to claim that human agents, unlike non-human agents, do not possess causal liabilities or the power to be caused to act in certain ways. But why should we accept this claim? His argument for this claim is that it supports genuine free will and genuine moral

responsibility. The problem with this argument is that it risks being circular. When asked why an agent-cause is uncaused, agent causalists would reply that it is necessary or required to ground or justify absolute free will. When asked why we have absolute free will, they would reply that it is necessary or required to ground or justify ultimate moral responsibility. When asked why we have ultimate moral responsibility, they would reply that it is because we have absolute free will. And when asked why we have absolute free will, they would reply that it is because an agent-cause is uncaused. Put simply, this argument risks begging the question by stipulating an uncaused agent-cause as a premise for the conclusion that there is absolute free will and ultimate moral responsibility, and by stipulating absolute free will and ultimate moral responsibility as premises for the conclusion that an agent-cause has to be uncaused. This argument assumes what it is supposed to prove. Independent arguments must be given for an uncaused agent-cause rather than assuming it just because it is a necessary condition of absolute free will and hence ultimate moral responsibility.

Second, I disagree with Lowe's claim that events of themselves possess no causal powers. On event-causal accounts, objects and substances exemplifying certain properties or being in certain states at certain times (or durations of time), and not of themselves, are what we describe as being magnetic, corrosive, inflammable, soluble, etc. And objects and substances having certain properties or being in certain states, and not of themselves, manifest or display their causal powers and liabilities, by acting on things, or being acted upon, in various appropriate ways, by attracting, corroding, burning, dissolving, etc. Substance causalists may question why we should prefer the event-causal account to the agent-causal account. Event-causalists can easily reply that their account can explain why, when, and how objects and substances manifest or display their causal powers and liabilities, by gaining or losing certain properties, or by changing to or persisting in certain states at certain times (or durations of

time). In comparison, substance causalists can hardly explain why, when, and how property-less and state-less objects and substances (of themselves) manifest or display their causal powers and liabilities by acting on things, or being acted upon by things. In short, event-causal accounts are empirically plausible and intelligible while substance-causal accounts are not. Since property-less and state-less objects and substances as causes lack empirical plausibility and intelligibility, it is reasonable to claim that objects and substances must have at least one property or be in at least one state at certain times (or durations of time) to qualify as causes. Moreover, given that an event-cause is defined as an object or a substance having a certain property or being in a certain state and thereby possessing certain causal powers at a certain time (or duration of time), and that an object or a substance always constitutes an event-cause together with their exemplification of properties, states, and powers at certain times (or durations of time), all causation is thus fundamentally event causation. The concept of a substance is then just a convenient abbreviation or abstraction of the concept of an event, and substance causal accounts are then just convenient abbreviations or abstractions of event causal accounts. Thus, it cannot be said that objects and substances are ontologically prior to events, and that event causation is reducible to substance causation.

Third, I find Lowe's claim that 'we use the language of event causation when we are at least partially ignorant about the causal agents that are at work' unconvincing because I do not see a significant difference between 'causal agents at work' or 'agents acting in certain manners' and 'objects having certain properties' or 'substances being in certain states'. They can be conceived as different descriptions of an event. Moreover, I find the opposing claim that 'we use the language of agent causation when we are at least partially ignorant about causal agents that are at work' convincing because objects or substances in general can be conceived as convenient abbreviations or abstractions of certain physical parts structured in certain ways

having certain properties, being in certain states, and thereby possessing certain causal powers and liabilities at certain times. Objects or substances need not be conceived as basic, primitive, simple, or un-analysable entities. Using Lowe's water example, water can be analysed as hydrogen and oxygen atoms structured in certain ways having certain properties and states like liquidity possessing the causal power to dissolve salt, which can be analysed as sodium and chlorine atoms structured in certain ways having certain properties or states like solidity possessing the causal liability to dissolve in water. It is not empirically plausible or intelligible to conceive water as a basic, primitive, simple, or un-analysable entity that can be property-less or state-less, for what else can water be if not a fusion of hydrogen and oxygen atoms having certain properties and states like liquidity. To say, in the un-analysed and coarse-grained language of agent causation that for water to dissolve salt is for water to cause salt to dissolve, is then just a convenient abbreviation or abstraction of the analysed and fine-grained language of event causation as described above. Mere use of abbreviated language or abstracted explanation does nothing to show that substances are basic, primitive, simple, or un-analysable entities, or that such entities of themselves can be causes.

Fourth, I agree with Lowe's plausible premise that we must be aware of ourselves as active causal agents before we can understand event causation and this involves having the capacity to move *ourselves* at will in accordance with our internal factors rather than being moved entirely by external factors. However, I disagree that this premise leads to the conclusion that acquiring the concept of agent causation is a pre-requisite of acquiring the concept of event causation, or that agent causation is conceptually prior to event causation. The analysis in the water example discussed above can be extended to human agents. Conceiving ourselves as active causal agents need not be conceiving ourselves as basic, primitive, simple or un-analysable substance causes. It can also be conceiving ourselves as event causes – as agents

having internal states like character and personality, desires and beliefs, preferences and values, and responsiveness and sensitivity to external circumstances and situations, at certain times (or durations of time). It is not empirically plausible or intelligible to conceive a human agent as a basic, primitive, simple, or un-analysable entity that can be property-less or state-less, for what else can an human agent be if not a fusion of organic molecules having certain psychological properties and states like character and personality, desires and beliefs, preferences and values, and responsiveness and sensitivity to external circumstances and situations, at certain times (or durations of time). On the event-causal view, what we acquire is the concept of event causation involving human agents rather than agent causation. Parallel to the agent-causal account, the event-causal account draws the different conclusion that event causation involving human agents is a pre-requisite of (or is conceptually prior to) event causation involving non-human agents. Acquisition of agent causation is then not a necessary pre-requisite of (or is conceptually prior to) event causation.

I believe that the problem is not that substances and substance-causes are un-analysable, but that there is a reluctance or even refusal on the part of the agent causalists to analyse substances and substance-causes. One motivation for this reluctance or refusal, in the case of human agents, is the worry that the causal relevance of agents seems to be lost in event-causal accounts, as causal relatas are between properties and states of the agent rather than the agent as a substance. The worry behind this motivation is exaggerated and unfounded, for on the event-causal account, the agent as a substance is a necessary constituent of (and hence are causally relevant to) an event or event-cause, by being the material and structural conditions of their properties and states. In the water example, water has the power of solvency by being hydrogen and oxygen atoms structured in certain ways having certain chemical properties. Analogously in the human agent example, the human agent has the

power to act in certain manners by being organic molecules structured in certain ways having certain biological and psychological properties. Agent causalists may rejoin by insisting that the worry remains because an event or event-cause involving a human agent is something happening to or something occurring in the agent, and hence the event or event-cause can never be an action or a doing. More pointedly, the human agent, as a necessary constituent of an event or event-cause, lacks the control required for that event or event-cause to count as an action or a doing. Drawing on Clarke's view on control above, my reply is that the human agent, as a necessary constituent of an event or event-cause, may possess both bare actional and rational control, whether or not causal determinism is true. And I believe that an event or event-cause involving a human agent with bare actional or rational control qualifies it as an action or a doing. Direct agential control that requires uncaused agent-causes is not required. If my arguments are acceptable, then contra Lowe, not only that event causation is not reducible to agent causation, but that agent causation is reducible to event causation. Here, agent causalists may press their rejoinder further by insisting that direct agential control or uncaused agent-cause is required to ground or justify absolute free will and hence ultimate moral responsibility. My reply is that this leads back to the risk of circularity objection introduced earlier. At this point, I hope to have shown why an agent-cause (whether as sole-cause or co-cause) as an explanation of action is inexplicable, unintelligible, and implausible (if not impossible) and there are adequate reasons to reject it.

**Agent Causation as Event Causation**

On the event causal account I favour, the human agent exemplifying intentional states – like desires, beliefs, and intentions – with goals, purposes, or reasons as their content or object as a single event-cause and the agent's action as a single event-effect. A human agent is a

persisting object or individual substance, with psychological capacities – like affection, cognition, and conation – that make intentional states possible, which in turn make the forming of attitudes and responses towards goals, purposes, and reasons as contents or objects possible. The human agent's psychological capacities, intentional states, as well as the reasons (content or object of intentional states) which he forms attitudes and responses towards are all indispensible components of agency and action. Both the agent himself (as a persisting object or individual substance with psychological capacities) and something about him (his intentional states and the reasons which he forms attitudes and responses toward) are causally relevant to his action. I believe that it is more intuitive to view agents as objects or substances possessing various psychological capacities with various mutually consensual and conflicting intentional states, who form attitudes and responses towards various facts of the situation that they happen to find themselves in, rather than just agents solely as objects and substances, as causes of their action. This view is consistent with there being antecedent causal conditions, whether they are sufficient or insufficient, deterministic or indeterministic.

Of course, my position is not without problems. To the agent causal theorists, my account faces the problems inherited from both causal determinism and causal indeterminism. Perhaps this is so. To be consistent with causal determinism, my account must say that there are sufficient causal conditions for an agent to exemplify his intentional states with reasons as their content. And to be consistent with causal indeterminism, my account must say that there may not be sufficient causal conditions for an agent to exemplify his intentional states with reasons as their content. Either way, my view is not able to ground or justify absolute free will and ultimate moral responsibility. My reply is that both causal determinism and causal indeterminism have not been proven true yet and I am not committed to either of them at this point. My proposal is only committed to the view that there are antecedent causal

conditions for an agent to exemplify his intentional states with reasons as their content and this view is consistent with both causal determinism and indeterminism. I believe that my event causal account is still preferable to the agent causal account which states that the agent as substance alone is the sole or co-cause of action, and that the agent's intentional states with reasons as their content are not causally relevant to (make no difference to, do not counterfactually support, do not contribute towards, or are not INUS conditions of) his action, where what causally explains his action ultimately falls back on some stipulated uncaused cause or unexplained explanans. Agent causal theorists may still insist that on my view, the agent's actions are caused and are not entirely up to him, and hence he cannot be held ultimately responsible for them. To this, my brief reply is to concede that the agent's action is indeed caused and are never entirely up to him, and hence he cannot be held ultimately responsible for them. But this does not imply that he is not morally responsible for them at all. Once the qualifiers 'entirely' and 'ultimately' are dropped, my view is compatible with the ascription of moral responsibility. The agent-causal account claims that since the ultimacy condition is required in our practice of moral responsibility ascription, entirety is required in our conception of the control the agent has over his action. And since entirety is required in our conception of the control the agent has over his action, uncaused agent-causes need to be postulated. In contrasts, my event-causal account claims that since the ultimacy condition is not required in our practice of moral responsibility ascription, entirety is not required in our conception of the control the agent has over his action. And since entirety is not required in our conception of the control the agent has over his action, uncaused agent-causes need not be postulated.

**References:**

Broad, C.D. (1952) 'Determinism, Indeterminism, and Libertarianism' in *Ethics and the History of Philosophy* (London RKP)

Chisholm, Roderick (1995) 'Agent, Causes, and Events: The Problem of Free Will' in *Agents, Causes, and Events* ed. Timothy O'Connor (Oxford University Press)

Chisholm, Roderick (1971) 'Reflection on Human Agency' in *Idealistic Studies I pp.40*

Clarke, Randolph (2010) 'Agent Causation' in *A Companion to the Philosophy of Action* ed. Timothy O'Connor and Constantine Sandis (Blackwell)

Clarke, Randolph (2009) from 'Agent Causation and Event Causation in the Production of Free Action' in *Free Will* ed. Derk Pereboom (Hackett)

Clarke, Randolph (2005) 'Toward a Credible Agent-Causal Account of Free Will' in *Agents, Causes, and Events* ed. Timothy O'Connor (Oxford University Press)

Clarke, Randolph (2003) *Libertarian Accounts of Free Will* (Oxford University Press)

Davidson, Donald (1997) 'Actions, Reasons, and Causes' in *The Philosophy of Action* ed. Alfred Mele (Oxford University Press)

Davidson, Donald (1980) 'Agency' in *Essays on Actions and Events* (Oxford University Press)

Ginet, Carl (1990) *On Action* (Cambridge University Press)

Goetz, S. (1988) 'A Non-Causal Theory of Agency' in *Philosophy and Phenomenological Research 49 pp. 303-316*

Kane, Robert (1995) 'Two Kinds of Incompatibilism' in *Agents, Causes, and Events* ed. Timothy O'Connor (Oxford University Press)

Kim, Jaegwon (1975) 'Events as Property Exemplifications' in *Action Theory* ed. Myles Brand & Douglas Walton (D. Reidel Publishing Company)

Lowe, E.J. (2008) *Personal Agency: The Metaphysics of Mind and Action* (Durham University Press)

Mackie, John L. (1993) 'Causes and Conditions' in *Causation* ed. Ernest Sosa & Michael Tooley (Oxford University Press)

O'Connor, Timothy (2009) 'Agent-Causal Power' in *Free Will* ed. Derk Pereboom (Hackett)

O'Connor, Timothy (2002) 'Libertarian Views: Dualist and Agent-Causal Theories' in *The Oxford Handbook of Free Will* ed. Robert Kane (Oxford University Press)

O'Connor, Timothy (2000) *Persons and Causes* (Oxford University Press)

O'Connor, Timothy (1995) 'Agent Causation' in *Agents, Causes, and Events* ed. Timothy O'Connor (Oxford University Press)

Shoemaker, Sydney (1984) 'Causality and Properties' in *Identity, Cause and Mind* (Cambridge University Press)

Strawson, Galen (1986) *Freedom and Belief* (Oxford University Press)

Taylor, Richard (1974) *Metaphysics 2nd Edition* (Prentice Hall)

Taylor, Richard (1966) *Action and Purpose* (Prentice Hall)

Walter, Henrik (2009) *Neurophilosophy of Free Will* (MIT Press)

# CHAPTER 4: MORAL RESPONSIBILITY WITHOUT ULTIMACY CONDITION

## Introduction

The issue of moral or legal responsibility is central to discussions on the freewill debate. It concerns our justification for moral praise and blame, or legal reward and punishment. A useful distinction can be drawn between retrospective (backward-looking/merit-based/desert-based) responsibility and prospective (forward-looking/consequence-based/efficacy-based) responsibility. The former looks backward on our intentions, commonly assumed to be a product of our deliberations under our control, which merits or deserves moral praise or blame, or legal reward and punishment. The latter looks forward on the outcome, consequences or efficacies of moral praise or blame, or legal reward and punishment on us.

There are a number of positions in the freewill debate with respect to moral responsibility. All incompatibilists hold that both freewill and moral responsibility are incompatible with causal determinism. They fall into three camps – Libertarians hold that causal determinism is false and that both freewill and moral responsibility are possible. Hard determinists hold that causal determinism is true and both freewill and moral responsibility are impossible. Hard incompatibilists remains uncommitted to the truth or falsity of causal determinism, but either way, both freewill and retrospective moral responsibility are impossible. All compatibilists hold that moral responsibility is compatible with causal determinism. Compatibilism falls into three camps – soft determinists hold that both freewill and moral responsibility are compatible with causal determinism and that causal determinism is true. Global controller theorists remain uncommitted to the truth or falsity of causal determinism, and hold that moral responsibility but not freewill is compatible with causal determinism. Reactive attitude

theorists remain uncommitted to the truth or falsity of causal determinism, hold that moral responsibility but not freewill is compatible with causal determinism, and that retrospective moral responsibility in the robust sense is possible. And since all positions agree with the possibility of prospective responsibility but disagree over the possibility of retrospective responsibility, the focus of this discussion will be on the disagreement – the retrospective aspect. Both hard and soft determinism usually focus solely on prospective responsibility and will not be discussed in this essay. The remaining positions will be evaluated in some detail. Hard incompatibilism argues for the impossibility of retrospective moral responsibility. Thomas Nagel (though not a hard incompatibilist) and Galen Strawson provide some of the strongest arguments for this position. Global controller theorists argue for the compatibility between retrospective moral responsibility and causal determinism in the absence of freewill (in the incompatibilist sense). Harry Frankfurt represents this position. Reactive attitude theorists argue for the irrelevance of causal determinism for moral responsibility. This position is represented by Peter Strawson. Libertarianism argues for the falsity of causal determinism (and hence affirm the freewill thesis) in order to make the robust sense of retrospective moral responsibility possible. This position is represented by Robert Kane.

My position, soft compatibilism, holds that the robust sense of retrospective moral responsibility is impossible no matter whether causal determinism or causal indeterminism is true, which is in line with the hard incompatibilist position. But this does not necessarily rule out a less robust sense of retrospective moral responsibility. The main idea is that the agent's possessing a certain character, personality, motivational structure (CPM), and a capacity for practical reasoning (CPR), who engages in a certain process of practical reasoning (exercising his CPR), at a certain time or duration of time (a type of event), is treated as a co-influent or co-determinant of other events; and is morally responsible only insofar as they are

co-influents or co-determinants. For this reason, choices and actions are not entirely up to the agent and do not have their origin solely in him, and so he cannot be morally responsible for them in the absolute or ultimate sense.

**Nagel on Moral Luck**

Thomas Nagel begins his paper 'Moral Luck' (Nagel 1979) by pointing out that "Kant believed that good or bad luck should influence neither our moral judgment of a person and his actions, nor his assessment of himself" (Nagel 1979, pp.24). On Kant's criterion of moral evaluation, all that is relevant is the agent's motivation, while situations, circumstances, character, or consequences are irrelevant. It clearly conforms to the intuition that "…people cannot be morally assessed for what is not their fault, or for what is due to factors beyond their control" (Nagel 1979, pp.25). Moral luck refers to the phenomenon where people are treated as objects of moral evaluation even when significant aspects of what they do depend on factors beyond their control. We intuitively believe that people should not be held morally responsible for what is beyond their control, and yet the more we examine people and their actions, the less we seem to be able to find factors which are within their control. So, when we start to take moral luck into consideration in our moral evaluation, it seems that there is nothing much people can be morally responsible for. Nagel maintains that the problem of moral luck arises from the nature of our moral evaluation itself. For if we subtract anything that merely happens from people's actions, we eventually discover that nothing remains after the subtraction. According to Nagel, there are four ways in which the natural objects of moral evaluation are subject to moral luck. The first is constitutive luck – the kind of person we are. This involves not only what we deliberately do, but our inclinations, capacities, and temperament as well. The second is circumstantial luck – the kind of problems and situations

we face. The third is causal luck – how our actions are determined by antecedent circumstances. And the fourth is resultant luck – what are the consequences of our actions. The main problem they pose is that when they are considered together, there is nothing much left for people to have control over. And it seems irrational to praise or blame for matters over which people have no control over, or for matters over which they have only partial control. Such matters may create the conditions for action, but action can be morally evaluated only to the extent that it goes beyond these conditions and that it does not just result from them. Note that moral luck is consistent with either causal determinism or causal indeterminism as some control over action is lost either way. Under causal determinism, there are some co-determinants of action people have no control over. And under causal indeterminism, there is some luck or random factors people have no control over.

Nagel recognises the obvious connection between the problem of moral luck, especially that of causal luck, and the problem of freewill. He says: "If one cannot be responsible for consequences of one's acts due to factors beyond one's control, or for antecedents of one's acts that are properties of temperament not subject to one's will, or for the circumstances that pose one's moral choices, then how can one be responsible even for the stripped-down acts of the will itself, if they are the product of antecedent circumstances outside of the will's control? The area of genuine agency, and therefore of legitimate moral judgment, seems to shrink under this scrutiny to an extensionless point. Everything seems to result from the combined influence of factors, antecedent and posterior to action, that are not within the agent's control. Since he cannot be responsible for them, he cannot be responsible for their results" (Nagel 1979, pp.35). If this is true, then it seems that our practice of ascribing moral responsibilities to agents is totally misconceived. Nagel suggests that this problem arises because the agent who acts and is the object of moral evaluation is reduced to a class of

events, that is, merely happenings. Moral evaluation of an agent is not an evaluation of what happens to the agent, but an evaluation of the agent. It is not an evaluation of a state of the world, or of an agent as part of the world. And it is not an evaluation of whether it would be better if the agent were different, or did not exist, or had not done some of the things the agent has done. We are evaluating the agent directly, not his existence or characteristics. The effect of focussing on the antecedent conditions that are not under the agent's control is to make this responsible agency disappear by reducing this object of moral evaluation into mere events. Nagel concludes that this problem has no solution because something in the idea of agency is incompatible with actions being events, or agents being objects. But once the external determinants of character, choices, and actions are gradually known, it becomes gradually clear that actions are events and agents are objects. As he remarked: "Eventually nothing remains which can be ascribed to the responsible self, and we are left with nothing but a portion of the larger sequence of events, which can be deplored or celebrated, but not blamed or praised" (Nagel 1979, pp.37).

Nagel ends by saying something about the internal view of agency: "There is a close connexion between our feelings about ourselves and our feelings about others. Guilt and indignation, shame and contempt, pride and admiration are internal and external sides of the same moral attitudes. We are unable to view ourselves simply as portions of the world, and from inside we have a rough idea of the boundary between what is us and what is not, what we do and what happens to us, what is our personality and what is an accidental handicap. We apply the same essentially internal conception of the self to others...We do not regard our actions and our characters merely as fortunate or unfortunate episodes – though they may also be that. We cannot simply take an external evaluative view of ourselves – of what we most essentially are and what we do. And this remains true even when we have seen that we are

not responsible for our own existence, or our nature, or the choices we have to make, or the circumstances that give our acts the consequences they have. Those acts remain ours and we remain ourselves, despite the persuasiveness of the reasons that seem to argue us out of existence" (Nagel 1979, pp.37). But the external view of events gradually erodes our internal view of agency through the gradual erosion of what we do by the subtraction of what happens.

While I accept that all four aspects of moral luck diminish moral responsibility to a large extent, I will examine the relationship between constitutive luck, causal luck, and agency to argue that the agent is not reduced to an 'extensionless point' and that moral responsibility is diminished but not entirely demolished. First, I view agency as a kind of event and the agent as a kind of object. Just as an event refers to an object (or a set of objects) exemplifying certain property (or properties) at certain times, an event can also refer to an agent (or a group of agents) exemplifying certain CPM at certain times (or durations to time). As the presence of CPM is a sufficient condition for an agent, the presence of an agent is a necessary (but insufficient) material and structural condition of his CPM. Perhaps the agent together with the relevant past and present circumstantial conditions provides jointly sufficient conditions of his CPM. So, there cannot possibly be any CPM that motivates choices and action without the presence of an agent. In other words, the presence of an agent is indispensible to any CPM that motivates choice and action. Second, the presence of the agent also suggests the presence of CPR, which contributes towards his choices and actions by engaging in the process of practical reasoning (exercising his CPR). So, the agent is not some passive entity that are simply carried along by antecedent conditions like CPM, but it is an active participant that brings his CPR to bear on his CPM by engaging in the process of practical reasoning (exercising his CPR). Even if it is true that his present exercising of CPR is conditioned by

his past CPM, his future CPM is in turn conditioned by his present exercising of CPR. So yes, agency may be a kind of event and an agent may be a kind of object. But agency is a type of event that involves an agent and an agent is a kind of object with CPM and CPR who engages in the process of practical reasoning. And yes, moral responsibility may be significantly diminished by the combined influences of the four aspects of moral luck. But assuming that an agent's CPM and CPR are functioning normally in the absence of coercion, compulsion, and constrain, this does not mean that the agent does not at least co-determine his choices and actions by engaging in the process of practical reasoning, and be morally responsible for them at least to that extent. Another way to state the point is this: given that the agent's choices and actions result from antecedent conditions which include his CPM, CPR, and the process of practical reasoning, I do not believe we can plausibly say that he is not at least partially responsible for them at all.

And I propose that the apparent conflict between the internal and external view is a combined result of our internal view of moral responsibility as an absolute or ultimate conception and our external view of events and things that fail to take into account of agency and agents as a constituent part of the natural scheme of things. Subtract 'what happens' from 'what we do' and we find nothing left because 'what we do' is part of 'what happens'. Acceptance of 'what we do' as part of 'what happens' implies the very presence of 'what we do', not its absence. I call this the 'internal within external' view. Although this view does reduce the role of agents themselves as causes of their choices and actions, it does not reduce them to 'extensionless points', as Nagel claims. Even on such a view, moral evaluation of an agent remains the evaluation of the agent's choices and actions in the world (internal within external), not merely an evaluation of what happens to the agent in the world (purely external), nor is it merely an evaluation of what is inside his head or within his mind, apart

from the world (purely internal). For the agent's choices and actions take place within the world and hence should be evaluated within it, and not apart from it. If we cannot accept that, it's only because we hold on to an absolute or ultimate conception of moral responsibility and elevate agency beyond the natural scheme of things.

**Galen Strawson on the Impossibility of Moral Responsibility**

In his paper 'The Impossibility of Moral Responsibility' (Strawson 2009a), Galen Strawson uses what he calls the Basic Argument to prove that we cannot be truly or ultimately morally responsible for our actions. According to this argument, we cannot be truly or ultimately morally responsible for our actions and whether causal determinism is true or false makes no difference. A short version of it can be given in three steps:

(1) Nothing can be causa sui - nothing can be the cause of itself.
(2) In order to be truly morally responsible for one's actions one would have to be causa sui, at least in certain crucial mental respects.
(3) Therefore nothing can be truly morally responsible.

Strawson restates a full version as follows:

(1) You do what you do because of the way you are.

So

(2) To be truly morally responsible for what you do you must be truly responsible for the way you are – at least in certain crucial mental respects.

But

(3) You cannot be truly responsible for the way you are, so you cannot be truly responsible for what you do.

Why can't you be truly responsible for the way you are? Because

(4) To be truly responsible for the way you are, you must have intentionally brought it about that you are the way you are, and this is impossible.

Why is it impossible? Well, suppose it is not. Suppose that

(5) You have somehow intentionally brought it about that you are the way you now are, and that you have brought this about in such a way that you can now be said to be truly responsible for being the way you are now.

For this to be true

(6) You must already have had a certain nature N in the light of which you intentionally brought it about that you are as you now are.

But then

(7) For it to be true you and you alone are truly responsible for how you now are, you must be truly responsible for having had the nature N in the light of which you intentionally brought it about that you are the way you now are.

So

(8) You must have intentionally brought it about that you had that nature N, in which case you must have existed already with a prior nature in the light of which you intentionally brought it about that you had the nature N in the light of which you intentionally brought it about that you are the way you now are...

Here, Strawson sets up a regress to show that nothing can be causa sui in the required way. He remarks: “We are what we are, and we cannot be thought to have made ourselves in such a way that we can be held to be free in our actions in such a way that we can be held to be

morally responsible for our actions in such a way that any punishment or reward for our actions is ultimately just or fair. Punishments and rewards may seem deeply appropriate or intrinsically 'fitting' to us in spite of this argument, and many of the various institutions of punishment and reward in human society appear to be practically indispensable in both their legal and non-legal forms. But if one takes the notion of justice that is central to our intellectual and cultural tradition seriously, then the evident consequence of the Basic Argument is that there is a fundamental sense in which no punishment or reward is ever ultimately just" (Strawson 2009a, pp.221).

As Strawson points out, step (2) of the restated Basic Argument can be rejected because the phrases 'truly responsible' and 'truly morally responsible' can be defined in many ways, and considers ways of doing so. The first is compatibilist. The compatiblists believe that we can be morally responsible even if causal determinism is true, just so long as we are not caused to act by any of a certain set of constraints, and that the act flows from our character or the way we are. So, compatibilist responsibility does not require that we should be truly responsible for how one is in any way at all, and hence step (2) is false. It is because we can have compatibilist responsibility even if the way we are is totally determined by factors entirely outside our control, compatibilist responsibility is often criticised for not amounting to any true moral responsibility in the robust sense. For if what we do is entirely because of the way we are, and we are in no way ultimately responsible for the way are, then it seems unjustified to praise or blame us for what we do. Strawson contends that even increasingly refined accounts of compatibilist responsibility can do nothing against this criticism. I disagree with Strawson here. Refined accounts of compatibilist responsibility need not affirm that the way we are is totally determined by factors outside our control. And moral responsibility (again, I am talking about the retrospective kind and not the prospective kind) need not be understood

in the true, robust, or ultimate sense, for it remains to be shown whether this true, robust, or ultimate sense of moral responsibility is a necessary requirement for our moral (praise and blame) and legal (reward and punishment) practices.

The second response is causal indeterminist libertarian, represented by Robert Kane. He argues that agents in an undetermined world can have freewill, for they can "have the power to make choices for which they have ultimate responsibility". That is, they can "have the power to make choices which can only and finally be explained in terms of their own wills (i.e. character, motives, and efforts of will)" (Kane 2009, pp.146). In claiming that we can be ultimately responsible for our present motives and character, Kane appears to accept step (2) and appears to accept that we have to 'make ourselves', and so be ultimately responsible for ourselves, in order to be morally responsible for what we do. Critics point out that the truth of causal indeterminism (or the falsity of causal determinism) does nothing to support moral responsibility. For the occurrence of partly luck or random factors cannot contribute in any way towards our being truly morally responsible for our action and character. If this is the case, then it seems that causal indeterminism cannot contribute in any way towards our justification of praise (reward) or blame (punishment).

The third response is the agent causal libertarian. It accepts that whether causal determinism is true or not, we cannot be held to be ultimately responsible for our character or personality or motivational structure. It then appeals to a certain picture of the self in order to argue that we can be truly free and morally responsible in spite of the fact that we cannot be held to be ultimately responsible for our character, personality or motivational structure. The argument can be set out as follows: we are free and truly morally responsible because our self (S) is independent of our character, personality or motivational structure (CPM). Our S considers our CPM and decides in the light of them, but our S possesses a causal power that is

independent of our CPM, in such a way that we can be truly morally responsible for our choices and actions, even though we are not truly responsible for our CPM. Step (2) is false because of the existence of S. Strawson then criticises this response as follows: S decides in the light of our CPM. But whatever S decides, either it decides as it does because of the way it is or because of some indeterministic factors in the decision process for which S cannot be responsible, and which cannot plausibly be thought to contribute towards S's true moral responsibility. And this brings us back to the Basic Argument. To be a source of true or ultimate responsibility, S must be responsible for being the way it is. But this is impossible, for nothing can be causa sui, and in order to be truly morally responsible for one's actions one would have to be causa sui, at least in certain crucial mental respects. As Strawson remarked: "The story of S and CPM adds another layer to the description of the human decision process, but it cannot change the fact that human beings cannot be ultimately self-determining in such a way as to be ultimately morally responsible for how they are, and thus for how they decide and act" (Strawson 2009a, pp.226).

Strawson's Basic Argument indeed provides powerful support for Nagel's constitutive luck. I agree with the general trust of the Basic Argument, which implies that true moral responsibility is impossible no matter whether causal determinism or causal indeterminism is true. And agent causation offers no help. I will question steps (1) to (3) of the Basic Argument for the Impossibility of Ultimate Responsibility (UR) in an attempt to limit constitutive luck but not eliminate it. In fact, I believe that the elimination of constitutive luck cannot be plausibly done because I find steps (4) to (8) of the Basic Argument for the Impossibility of Causa Sui (CS) sound. To appreciate steps (1) to (3) of the Basic Argument better, let's simplify them into a Modus Tollens: if UR is possible, then CS is possible. But CS is impossible. Hence, UR is impossible. Or more simply, UR implies CS. CS is false.

Hence, UR is false. So, the most powerful premise of the Basic Argument is that CS is impossible or false. Now, to appreciate steps (4) to (8) of the Basic Argument in support for this powerful premise better, let's simplify them into another Modus Tollens: if it is possible for S (self) to be CS (the cause of itself), then it must be possible for S to exist prior to the act of CS (causing itself). But it is impossible for S to exist prior to the act of CS. Hence, it is impossible for S to be CS. Or more simply, 'S can be CS' implies 'S exists prior to the act of CS'. But it cannot be true that 'S exists prior to the act of CS'. Hence, it cannot be true that 'S can be CS'. This is the logical problem of affirming CS, for the claim 'S can be CS' entails that 'S can exist prior to the act of CS' and yet 'S cannot exist prior to the act of CS', and this generates a contradiction. Steps (4) to (8) are sound and they provide strong support for step (1). Let's look at steps (2) and (3) next.

As Strawson admits that step (2) of Basic Argument can be rejected because qualifiers like 'truly' and 'ultimately' can be defined in many ways, it would be useful to fix their definitions before discussing my response to steps (1) to (3) of the Basic Argument. Reading 'truly' or 'ultimately' in step (2) as 'solely' or 'completely', we can reformulate step (2) in the short version as follows: in order to be 'solely' and 'completely' morally responsible for one's actions one would have to be causa sui, at least in certain crucial mental respects; and in the full version as follows: to be 'solely' or 'completely' morally responsible for what you do you must be 'solely' or 'completely' responsible for the way you are – at least in certain crucial mental respects. Thus read, step (2) seems perfectly acceptable. And reading 'truly' or 'ultimately' in step (3) as 'solely' or 'completely', we can reformulate step (3) in the short version as follows: nothing can be 'solely' and 'completely' morally responsible; and in the full version as follows: you are not 'solely' or 'completely' morally responsible for the way

you are – at least in certain crucial mental respects, so you are not 'solely' or 'completely' responsible what you do. Thus read, step (3) seems perfectly acceptable too.

My view on steps (1) to (3) of the full version is as follow. My first point is to treat the agent S and his possession of CPM as an object-property relation typifying a simple event. S can be conceived of as an object (or substance). CPM can be conceived as the properties of S. As properties in general can be conceived as the ways objects are, CPM can be conceived of as the way S is. S possessing a certain CPM at a certain time or duration of time can be conceived as a simple event. S is a necessary (but insufficient) material and structural condition of CPM, and CPM is a sufficient condition of S. Perhaps S together with the relevant past and present circumstantial conditions provides the jointly sufficient conditions for CPM. The presence of S is indispensible to CPM that motivates choice and action. My second point is to accept that S, on top of CPM, possesses CPR which enables the process of practical reasoning. So, S is not some passive entity that is simply carried along by antecedent conditions like CPM, but he is an active participant that brings his CPR to bear on his CPM by engaging in the process of practical reasoning (exercising his CPR).

What implications do my two points have for step (1) to (3)? Step (1) says that S's CPM causes his action A. Step (2) says that S has to be responsible for his CPM in order to be responsible for his A. Step (3) says that S cannot be responsible for his CPM and hence cannot be responsible for his A. On step (1), I would say that it is true but incomplete as it leaves something crucial out, namely the exercising of CPR. S's exercising of his CPR in the present is conditioned by his past CPM, and his future CPM is in turn conditioned by the exercising of his CPR in the present. Of course, S has to reach a certain maturity before his CPR is adequately developed. Moreover, S already has a certain CPM before his CPR is adequately developed. But once S's CPR is adequately developed, he is able to influence his

CPM by exercising his CPR. To complete the picture, as long as S has an adequately developed CPR in the present, S is able to condition his future CPM by exercising his CPR in the present (in the absence of coercion, compulsion, and constrain). Carrying this line of thought further, as long as S has an adequately developed CPR in the past, S is able to condition his present CPM by exercising his CPR in the past (in the absence of coercion, compulsion, and constrain). And this stretches all the way back until S's CPR has not been adequately developed yet. If this is true, then its implication for step (2) is this: S is partially but not 'truly', 'ultimately', 'solely', or 'completely' responsible for the way he is. This is because S is responsible for his CPM only after his CPR is adequately developed, not before. And if S is partially morally responsible for the way he is, then he is partially morally responsible for what he does. Note that steps (2) and (3) remain true despite my two points.

My third point is then to accept that steps (2) and (3) are true but say that they do not rule out the possibility that we are at least partially morally responsible. More specifically, we do not need qualifiers like 'truly', 'ultimately', 'solely', 'completely' in the requirement for partial moral responsibility. In reply to the short version of step (2), we do not need to be causa sui in order for us to be partially morally responsible for our action. And in reply to the full version, we do not need to be 'truly', 'ultimately', 'solely', or 'completely' morally responsible for the way we are in order to be partially morally responsible for our actions. In short, I affirm the possibility of moral responsibility but reject the absolute or ultimate conception of it. While Strawson concludes that both causal determinism and causal indeterminism rule out moral responsibility, I conclude that both causal determinism and causal indeterminism rule out moral responsibility in the absolute and ultimate sense, but not in the partial sense. This result of my argument is significant because it shows that both

causal determinism and causal indeterminism can be made consistent with moral responsibility, but not the ultimacy condition.

**Frankfurt on Alternative Possibilities & Moral Responsibility**

Harry Frankurt argues against the principle of alternative possibilities (PAP) in his essay 'Alternate Possibilities and Moral Responsibility' (Frankfurt 2009). This principle states that a person is morally responsible for what he has done only if he could have done otherwise. This principle is false according to Frankfurt. A person may well be morally responsible for what he has done even though he could not have done otherwise. Frankfurt illustrates the falsity via an example (the first among many others now famously known as Frankfurt-style examples):

"Suppose someone, Black, let us say, wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something other than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones initial preferences and inclinations, then, Black will have his way. What steps will Black take, if he believes he must take steps, in order to ensure that Jones decides and acts as he wishes? Anyone with a theory concerning what "could have done otherwise" means may answer this question for himself by describing whatever measures he would regard as sufficient to guarantee that, in the relevant sense, Jones cannot do otherwise. Let Black pronounce a

terrible threat, and in this way both force Jones to perform the desired action and prevent him from performing a forbidden one. Let Black give Jones a potion, or put him under hypnosis and in some such way as these generate in Jones an irresistible inner compulsion to perform the act Black wants performed and to avoid others. Or let Black manipulate the minute processes of Jone's brain and nervous system in some more direct way, so that causal forces running in and out of his synapses and along the poor man's nerves determine that he chooses to act and that he does act in the one way and not in any other. Given any conditions under which it will be maintained that Jones cannot do otherwise, in other words, let Black bring it about that those conditions prevail. The structure of the example is flexible enough, I think, to find a way around any charge of irrelevance by accommodating the doctrine on which the charge is based" (Frankfurt 2009, pp.172-3).

"Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. It would be quite unreasonable to excuse Jones for his action, or to withhold the praise to which it would normally entitle him, on the basis of the fact that he could not have done otherwise. This fact played no role at all in leading him to act as he did. He would have acted the same even if it had not been a fact. Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it. In this example there are sufficient conditions for Jone's performing the action in question. What action he performs is not up to him. Of course it is in a way up to him whether he acts on his own or as a result of Black's intervention. That depends upon what action he himself is inclined to perform. But whether he finally acts on his own or as a result of Black's

intervention, he performs the same action. He has no alternative but to do what Black wants him to do. If he does it on his own, however, his moral responsibility for doing it is not affected by the fact that Black was lurking in the background with sinister intent, since this intent never comes into play" (Frankfurt 2009, pp.173-4).

The most important moral Frankfurt wants to draw from his example is this: "Even though the person was unable to do otherwise, that is to say, it may not be the case that he acted as he did because he could not have done otherwise. Now if someone had no alternative to performing a certain action but did not perform it because he was unable to do otherwise, then he would have performed exactly the same action even if he could have done otherwise. The circumstances that made it impossible for him to do otherwise could have been subtracted from the situation without affecting what happened or why it happened in any way. Whatever it was that actually led the person to do what he did, or that made him do it, would have led him to do it or made him do it even if it had been possible for him to do something else instead. Thus it would have made no difference, so far as concerns his action or how he came to perform it, if the circumstances that made it impossible for him to avoid performing it had not prevailed. The fact that he could not have done otherwise clearly provides no basis for supposing that he might have done otherwise if he had been able to do so. When a fact is in this way irrelevant to the problem of accounting for a person's action it seems quite gratuitous to assign it any weight in the assessment of his moral responsibility. Why this fact should be considered in reaching a moral judgment concerning the person when it does not help in any way to understand either what made him act as he did or what, in other circumstances, he might have done?" (Frankfurt 2009, pp.174)

For Frankfurt, PAP is mistaken because it asserts that a person bears no moral responsibility – that is, he is to be excused – for having performed an action if there were circumstances

that made it impossible for him to avoid performing it. But there may be circumstances that make it impossible for a person to avoid performing some action without those circumstances in any way bringing it about that he performs that action. A person cannot appeal to circumstances of this sort to absolve himself of moral responsibility for performing an action. For those circumstances, by hypothesis, actually had nothing to do with his having done what he did. He would have done precisely the same thing, even if they had not prevailed. It may be true that there were circumstances that made it impossible for a person to avoid doing something, and that these circumstances actually played a role in bringing it about that he did it, so that it is correct to say that he did it because he could not have done otherwise. But the person really wanted to do what he did, and he did it because it was what he really wanted to do, so that it is not correct to say that he did what he did only because he could not have done otherwise. Under these conditions, the person may well be morally responsible for what he has done. On the other hand, he will not be morally responsible for what he has done if he did it only because he could not have done otherwise, even if what he did was something he really wanted to do.

One way to understand Frankfurt's view is to ask 'would a person have done otherwise even if he could at the time when he acts, given identical past conditions and circumstances?' I take Frankfurt's answer to be 'he would not if what he did was what he really wanted at the time when he acts, given identical past conditions and circumstances'. And if a person would have done otherwise if he could at the time when he acts, then what he did was not what he really wanted. When it comes to the ascription of moral responsibility, what is required is what the person really wanted to do such that he would not have done otherwise even if he could at the time when he acts, given identical past conditions and circumstances. The past conditions and circumstances include the person's CPM, CPR, process of practical reasoning,

as well as the desires, beliefs, intentions, and reasons that figure in the process of practical reasoning. And the person's desires and beliefs include his preferences and values, as well as arguments and evidences available to him at the time when he acts. Whether or not he could have done otherwise at the time when he acts is not required for the ascription of moral responsibility. If a person claims that 'I would have done otherwise if I could', his claim already implies the benefit of hindsight. With the benefit of hindsight, the arguments and evidences available to him, his preferences and values, or desires and beliefs, could have changed sometime after he had acted. That is, with the benefit of hindsight, the conditions and circumstances differ from the time before he had acted. Denial of PAP requires denying the stronger claim that a person would have done otherwise even if he could at the time when he acts, given identical past conditions and circumstances. It is does not require denying the weaker claim that a person would have done otherwise even if he could at the time when he acts, given similar (but not identical) past conditions and circumstances. It is worth noting that Frankfurt need not deny that there are alternative possibilities, he need only to deny that even if there are alternative possibilities, it is not required for moral responsibility ascription.

There are many objections to Frankfurt's view, with the most prominent being the 'What-should-he-have-done defence' or 'W-defence' made by several libertarians. A modified version from David Widerker in his paper 'Responsibility and Frankfurt-type Examples' goes like this: "I understand that you, Frankfurt, want to hold Jones (in the above example) morally responsible for doing what he did. If so, tell me what, in your opinion, should Jones have done instead? Now, you cannot say that Jones should have done otherwise because this is not within his power to do. Hence, I do not see how you can hold Jones responsible for doing what he did" (Widerker 2002, pp.329). This objection suggests that PAP is a necessary

condition for ascribing moral responsibility. And there are many responses to this objection as well as rejoinders to these responses but I shall not be discussing them here.

The following is my take on Frankfurt's view and the 'W-defence' and their implication for moral responsibility ascription. First, focussing too much on the Frankfurt-style examples may lead one to mistake manipulation for causal determinism. There are hard incompatibilists who argue that manipulation is really no different from causal determinism with respect to moral responsibility ascription but I will not discuss it here. Here, I assume that there is a significant difference between manipulation and causal determinism, where the former requires the presence of a will or design while the latter does not. In these examples, Black plays the role of the 'global controller' or 'counterfactual intervener' that manipulates Jones' behaviour in a way that he cannot perform any act forbidden by Black. Now, if Jones chose to perform a forbidden act, Black can intervene to bring it about that he does not perform that forbidden act. The 'W-defence' seems right in this case that Jones is not morally responsible for not choosing to perform the forbidden act, Black is. But for Frankfurt however, Black's role is to eliminate alternative possibilities, and not interfere with what Jones really wanted to do. Moreover, Frankfurt has made it clear that he does not need Black or any other manipulator to make his point. For the 'manipulator' role can be played by any natural forces involving no will or design at all. What Frankfurt has in mind is actually causal determinism, not manipulation. So, the alternative sequence in the Frankfurt-style examples can be dispensed with in favour of the actual sequence. That is, focus on causal determinism (what Frankfurt intended) and not manipulation (what Frankfurt did not intend). Second, these natural deterministic forces can be broad enough to encompass both external and internal determinants. And the internal determinants here include Jones' CPM,

CPR, process of practical reasoning, as well as the desires, beliefs, intentions, and reasons that figure in the process of practical reasoning.

The crux of the issue is this: should we hold Jones morally responsible if his action is at least co-determined by his CPM, CPR, process of practical reasoning, as well as the desires, beliefs, intentions, and reasons that figure in the process of practical reasoning, such that he could not have done otherwise? Or should we hold Jones morally responsible only if he could have done otherwise, which implies that his action is not at least co-determined by his CPM, CPR, process of practical reasoning, as well as the desires, beliefs, intentions, and reasons that figure in the process of practical reasoning? While Frankfurt answers in the affirmative to the first question and in the negative to the second, supporters of the 'W-defence' answer in the affirmative to the second question and in the negative to the first. What Frankfurt is trying to say is that we are holding Jones morally responsible for his CPM, CPR, process of practical reasoning, as well as the desires, beliefs, intentions, and reasons that figure in the process of practical reasoning, leading to his choice and action, and not because he could have done otherwise. In other words, we are holding Jones morally responsible because Jones' choice and action follow from his internal determinants, and not because Jones' choice and action may or may not follow from his internal determinants. And since we are not holding Jones morally responsible because he could have done otherwise, ascriptions of moral responsibility is compatible with causal determinism. Supporters of the 'W-defence' may insist that Jones' choice and action is not really his if he could not have chosen or acted otherwise. To this, Frankfurt would have asked how can we say that Jones' choice or action is not really his if he has chosen and acted on what he really wanted, or what follows from his internal determinants, or what follows from his CPM, CPR, process of

practical reasoning, as well as the desires, beliefs, intentions, and reasons that figure in the process of practical reasoning.

The advantage of Frankfurt's position is that it does not need to affirm either causal determinism or causal indeterminism because moral responsibility can be preserved either way. I want to add that Frankfurt's position can be made more plausible by dropping the absolute or ultimate conception of moral responsibility. Since there are external and internal co-determinants not within Jones' control, his choice and action are not entirely up to him or have their origin solely in him. Consequently, it is more plausible to say that Jones' choice and action is only partly his, rather than wholly his. The disadvantage of the 'W-defence' is that affirmation of PAP requires the truth of causal indeterminism and hence it cannot be used to support the absolute or ultimate conception of moral responsibility. This is because if causal indeterminism is true, then Jones' internal determinants like CPM, CPR, process of practical reasoning, as well as the desires, beliefs, intentions, and reasons that figure in the process of practical reasoning, do not co-determine his choice and action, for they are always subject to alternative possibilities. On causal determinism, the conjunction of relevant internal and external determinants provides sufficient causal conditions for Jones' choice and action. And it adopts a weaker notion of conditional alternative possibilities that imply some difference in any of the component of internal and external determinants. On causal indeterminism, the conjunction of relevant internal and external determinants does not provide sufficient causal conditions for Jones' choice and action. And it adopts a stronger notion of unconditional alternative possibilities that imply no difference in any of the component of internal and external determinants. But if the conjunction of Jones' internal and external determinants neither provides sufficient causal conditions of nor account for the differences in Jones' choice and action, then what causally explains the alternative

possibilities seems to be luck or random factors. And since Jones has no control over luck or random factors, his choice and action is not entirely up to him or have their origin solely in him. And if Jones' choice and action are not entirely up to him or have their origin solely in him, then it seems that we cannot plausibly ascribe a more robust sense of moral responsibility to Jones. Further premises (perhaps agent causation), on top of causal indeterminism, are required to justify a more robust sense of moral responsibility desired by the supporters of 'W-defence'. However, I believe that the 'W-defence' can be made more plausible if it affirms causal indeterminism along with a less robust sense of moral responsibility. Supporters of the 'W-defence' can then claim that although the conjunction of Jones' internal and external determinants does not causally determine his choice and action, but it does make a difference to or contribute towards them. This only supports a less robust sense of moral responsibility. And I believe that both causal determinism and causal indeterminism are compatible with a less robust sense of moral responsibility, but not with the more robust one.

**Peter Strawson on Participant Reactive Attitudes**

Peter Strawson's aim in his paper 'Freedom and Resentment' (Strawson 2009), is an analysis of what he calls participant reactive attitudes and the effect of accepting the truth of a general thesis of causal determinism has upon them. Strawson defines participant reactive attitudes as "essentially natural human reactions to the good or ill will or indifference of others towards us, as displayed in their attitudes and actions" (Strawson 2009, pp.80). And he wishes to examine whether accepting the truth of causal determinism leads to the decay or the repudiation of all such participant reactive attitudes. He thinks that the important question here is what it would be rational to do if causal determinism were true, a question about the

rational justification of participant reactive attitudes in general, and that those who take this question seriously do so only because they utterly failed to appreciate our natural human commitment to these ordinary interpersonal attitudes. To emphasise the entrenchment of such ordinary interpersonal attitudes in our lives, he says: "This commitment is part of the general framework of human life, not something that can come up for review as particular cases can come up for review within this general framework" (Strawson 2009, pp.83). And if we could imagine that we have no choice over our commitment to these ordinary interpersonal attitudes, then we could choose rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of causal determinism would have no bearing on the rationality of this choice.

Strawson then goes on to compare two positions on the effect of accepting the truth of the thesis of causal determinism has upon our participant reactive attitudes – the pessimist and the optimist. They roughly correspond to the libertarian and the compatibilist positions respectively. Basically, the pessimist claims that our participant reactive attitudes, which refer more or less to our retrospective responsibility, can never be justified if causal determinism is true. The optimist claim that the truth of causal determinism would not shake the foundations of the concept of moral responsibility and of the practices of moral praise and blame, which refers more or less to prospective responsibility or the efficacy of these practices in regulating behaviour in socially desirable ways. These practices exemplify what Strawson calls thoroughgoing objectivity of attitude as they are represented solely as instruments of policy, as methods of individual treatment and social control. The pessimist recoils emotionally from this picture because a thoroughgoing objectivity of attitude, which excludes the participant reactive attitudes, excludes at the same time essential elements in the

concepts of (retrospective) moral responsibility. However, this emotional shock is a reaction, not simply to an inadequate concept of moral responsibility, but to the optimist's suggestion in cultivating an exclusive objectivity of attitude in all moral cases, and to set aside the ordinary interpersonal attitudes.

Strawson criticises both the optimist and pessimist for over-intellectualising the facts in different ways. In defending the independence of participant reactive attitudes from theoretical convictions, P. Strawson remarks: "Inside the general structure or web of human attitudes and feelings, there is endless room for modification, redirection, criticism, and justification. But questions of justification are internal to the structure or relate to modifications internal to it. The existence of the general framework of attitudes itself is something we are given with the fact of human society. As a whole, it neither calls for, nor permits, an external 'rational' justification" (Strawson 2009, pp.91). Both the pessimist and optimist are unable to accept this in different ways. Here is P. Strawson's evaluation of their positions: "The optimist's style of over-intellectualizing the facts is that of a characteristically incomplete empiricism, a one-eyed utilitarianism. He seeks to find an adequate basis for certain social practices in calculated consequences, and loses sight (perhaps wishes to lose sight) of the human attitudes of which these practices are, in part, the expression. The pessimist does not lose sight of these attitudes, but is unable to accept the fact that it is just these attitudes themselves which fill the gap in the optimist's account. Because of this, he thinks the gap can be filled only if some general metaphysical proposition is repeatedly verified, verified in all cases where it is appropriate to attribute moral responsibility. This proposition he finds it as difficult to state coherently and with intelligible relevance as its determinist contradictory. Even when a formula has been found ('contra-causal freedom' or

something of the kind) there still seems to remain a gap between its applicability in particular cases and its supposed moral consequences" (Strawson 2009, pp.92).

However, Strawson believes that the view of the optimist is only right if sufficiently modified. And he concludes: "It is far from wrong to emphasize the efficacy of all those practices which express or manifest our moral attitudes, in regulating behaviour in ways considered desirable; or to add that when certain of our beliefs about the efficacy of some of these practices turn out to be false, then we may have good reason for dropping or modifying those practices. What is wrong is to forget that these practices, and their reception, the reactions to them, really are expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes. Our practices do not merely exploit our natures, they express them. Indeed the very understanding of the kind of efficacy these expressions of our attitudes have turns on our remembering this. When we do remember this, and modify the optimist's position accordingly, we simultaneously correct its conceptual deficiencies and ward off the dangers it seems to entail, without recourse to the obscure and panicky metaphysics of libertarianism" (Strawson 2009, pp.93).

Strawson's view seems very attractive in many ways. First, he affirms participant reactive attitudes, which are essential to the concept of retrospective moral responsibility. For it seems to be true that we have deeply entrenched natural commitments to our participant reactive attitudes, to the extent that even if the thesis of causal determinism turns out to be true, these ordinary interpersonal attitudes seems indispensible to the way we normally relate to one another. Second, he rejects both the pessimist (libertarian) and optimist (compatibilist) positions for over-intellectualising in different directions. Although he agrees with the pessimist in affirming participant reactive attitudes, he disagrees with the pessimist on the need to justify our participant reactive attitudes with general metaphysical propositions like

the freewill thesis, which is not only difficult to state coherently but also as irrelevant as its opposing causal determinism thesis. And although he agrees with the optimist on the usefulness of prospective responsibility based on the efficacy of such practices, he disagrees with the optimist's acceptance of the causal determinism thesis and the resultant rejection of retrospective responsibility based on our ordinary interpersonal attitudes, practices which seem to be an indispensible part of the general framework of human life. So, Strawson's view is compatibilist only insofar as moral responsibility is compatible with causal determinism because causal determinism is irrelevant to moral responsibility even if it is true. It affirms retrospective moral responsibility in the robust sense while remains uncommitted to the truth or falsity of causal determinism.

I agree with Strawson that our participant reactive attitudes are too deeply entrenched to be given up no matter whether causal determinism or causal indeterminism is true. But I believe that the question of justification for our participant reactive attitudes remains to be answered, and answering this question requires us to confront the theses of causal determinism and causal indeterminism rather than avoid them. Strawson's claim that questions about the general structure or web of human attitudes and feelings can only have justifications internal to the structure and they neither call for nor permit external 'rational' justifications needs further support. The only reason he gives for its acceptance is that the existence of this general framework of attitudes itself is something we are given with the fact of human society. Both the pessimist and optimist supporters of external 'rational' justifications can respond in a number of ways. Consider the 'W-defence' as discussed above. If causal determinism were true, Jones could not have done otherwise. And if Jones could not have done otherwise, what else should Jones have done? So are we or are we not justified to apply our participant reactive attitude to Jones? If we base it on an external 'rational' justification,

the thesis of causal determinism in this case, it seems that we are not justified to do so. Both the pessimist and optimist would agree that the 'W-defence' is more than just a theoretical conviction, for it does have practical implications. So it follows that the truth or falsity of causal determinism does have an effect on our participant reactive attitudes. In accepting the 'W-defence', the pessimist would have to invent 'obscure and panicky' metaphysics to save participant reactive attitudes (and in turn retrospective moral responsibility) and the optimist would have to give up participant reactive attitudes (and in turn retrospective moral responsibility) while retaining prospective moral responsibility. What if despite knowing this, we simply ignore this external 'rational' justification and embrace only internal justifications of the general framework of attitudes? The result would reduce the general framework of attitudes to a mere social construct formed by our shared customs and habits (fact of human society). Or worse, acceptance of a plausible external 'rational' justification may render this general framework of attitudes a false hypothesis or an error theory, to be revised or repudiated. My view is that there may be a way to save participant reactive attitudes, but not simply by prioritising our shared customs and habits (fact of human society) over some of our better established theoretical convictions (like causal determinism or causal indeterminism). And in confronting our better established theoretical convictions, our participant reactive attitudes may not survive intact, and may need to be revised.

**Kane on Ultimate Responsibility**

Robert Kane, in his book *The Significance of Free Will* (Kane 1996), defines free will as "the power of agents to be the ultimate creators (or originators) and sustainers of their own ends and purposes" (Kane 1996, pp.4). It is to be distinguished from freedom of action. Freedom of action unhindered in the pursuit of your purposes and it is compatible with the thesis of

causal determinism. Freedom to will is to be the ultimate source or creator of your own purposes and it is incompatible with the thesis of causal determinism. There are two conditions for the freedom to will. The first is the alternative possibilities (AP) condition – the requirement that the free agent could have willed otherwise. The second condition, ultimate responsibility, accounts for the "ultimate" in the original definition of free will: "the power of agents to be the *ultimate* creators and sustainers of their own ends or purposes." Freedom to will (but not freedom to act) is about the forming and shaping of character traits and motives that are the sources or origins of praiseworthy or blameworthy choices and actions. But to be ultimately responsible for the agent's choices and actions, he or she must be responsible for anything that is a sufficient reason (condition, cause or motive) for the action's occurring. The relation between AP and UR us this: UR does not require AP for every act done of our own free wills, but it does require AP with respect to some acts in our past life histories by which we formed our present character traits. In other words, UR does not rule out the possibility that our choices and actions may be determined by our wills, character traits, and motives, but it does require that whenever this is so, to be ultimately responsible for the way we are, we must be responsible for forming the wills, character traits, and motives that eventually determine our choices and actions. And the key difference between AP and UR is this: while the AP condition focuses on notions like 'necessity', possibility', 'power', 'ability', 'can', and 'could have done otherwise', the UR condition focuses on notions like 'sources', grounds', 'reasons', and 'explanations' of our wills, character traits, and motives. Following Aristotle, Kane suggests that to have freewill, our choices and actions must be 'up to us'. And the concept of our choices and actions being 'up to us' is connected with the concept that the origin (source, ground, reason, explanation) of our choices and actions is 'in us'.

In criticising Kane's concept of UR, Galen Strawson agrees that UR requires a kind of self-origination, and characterises ultimate moral responsibility by reference to the story of heaven and hell. UR is responsibility of such a kind that, if we have it, it makes sense to propose that it could be just to punish some of us with torment in hell and reward others with bliss in heaven. It makes sense because what we do can be absolutely up to us. He presents the argument that UR requires radical self-creation in two steps:

(1) When you act, you do what you do, in the situation in which you find yourself, because of the way you are.
(2) So if you are to be UR for what you do, you must somehow or other be UR for the way you are-at least in certain crucial mental respects.

Strawson then argues for the impossibility of UR, reminiscent of his Basic Argument as discussed above, as follows:

(3) If you are to be UR for the way you are, you must have intentionally brought it about that you are the way you are.

And the problem is then this. Suppose

(4) You have somehow intentionally brought it about that you are the way you now are, in certain mental respects: suppose you have brought it about that you have a certain mental nature Z in such a way that you can now be said to be UR for Z.

For this to be true

(5) You must already have had a mental nature Y, in the light of which you intentionally brought it about that you now have nature Z. (If you didn't already have a mental

nature then you didn't have any intentions or preferences and can't be UR for the way you now are, even if you have changed.)

But then

(6) For it to be true that you are UR for how you now are, you must be UR for that nature, Y, in the light of which you brought it about that you now have nature Z.

So

(7) You must have intentionally brought it about that you had Y.

But then

(8) You must have existed already with a prior nature, X, in the light of which you brought it about that you had Y, in the light of which you brought it about that you now have Z.

Here, a potentially infinite regress occurs. Strawson argues that ultimate self-origination is the only thing that could actually ground UR. In order for one to be truly or ultimately responsible for how one is, in such a way that one can be truly morally responsible for what one does, something impossible has to be true: there has to be, and cannot be, a starting point in the series of acts of bringing it about that one has a certain nature; a starting point that constitutes an act of ultimate self-origination. This argument is completely independent of whether causal determinism is true or false. We cannot have UR if the antecedent conditions of our choices or actions are causally indeterministic (that is, subject to luck or random factors). Hence, Strawson concludes that both causal determinism and causal indeterminism are incompatible with UR. I believe that Strawson's attack on UR is devastating and conclusive. His main point can be summarised in the short version of his Basic Argument, (1) nothing can be causa sui (the cause of itself), (2) UR requires something to be causa sui, and (3) hence, nothing can be UR. The conclusion turns out true no matter whether causal

determinism or causal indeterminism is true. And I believe that the only way to preserve moral responsibility is to separate U (ultimacy condition) from R (moral responsibility), and then to reject U and accept R. That is, a more robust sense of moral responsibility has to be given up for a less robust one, on pain of facing Strawson's Basic Argument.

Kane may object that a less robust sense of moral responsibility is no longer the traditional sense of moral responsibility as we understand it, and that this latter sense of moral responsibility is worth defending. I disagree. First, I have already made clear that by a less robust sense of moral responsibility, I mean the retrospective aspect (what is under threat and what libertarians want to defend) and not the prospective aspect (what is acceptable to all positions). Second, must all retrospective moral responsibility be understood in the absolute or 'heaven and hell' sense? If not, why do we need the ultimacy condition for? Both Kane and Strawson agree that if we are ultimately responsible for our choices and actions, then we are ultimately responsible for our wills, character traits, and motives. While Kane argues via modus ponens that the antecedent is true and hence the consequent must be true, Strawson argues via modus tollens that the consequent is false and hence the antecedent must be false. I agree with Strawson, but would like to add that we do not need the qualifier 'ultimately' in this argument. That is, I accept the following argument: if we are (partially) responsible for our choices and actions, then we are (partially) responsible for our CPM. We are (partially) responsible for our choices and actions. Hence, we are (partially) responsible for our CPM. I have shown in my discussion of Strawson above that being partially responsible for our CPM is possible as long as we have an adequately developed CPR. Put simply, I accept R but reject U. Kane wants to say that our choices and actions are <u>entirely</u> 'up to us' only when they have their origin <u>solely</u> 'in us'. This is what the ultimacy condition requires. In the face of Nagel's and Strawson's arguments, I believe that we can cast doubt on the ultimacy

qualifiers ‘entirely’ and ‘solely’. To say that our choices and actions are ‘up to us’ only when they have their origin ‘in us’, while leaving out the ultimacy qualifiers ‘entirely’ and ‘solely’, sounds much more plausible to me. This is the view that our choices and actions are ‘up to us’ (but not entirely) and have their origin ‘in us’ (but not solely) when they follow from our CPM, CPR, process of practical reasoning as well as the desires, beliefs, intentions, and reasons that figure in the process of practical reasoning, and in the absence of coercion, compulsion, and constraint. This view in no way implies that our choices and actions are not ‘up to us’ or do not have their origin ‘in us’ at all. Third, does our actual practice of ascribing moral responsibility necessarily involve the ultimacy condition? I believe that it does not. In our moral practice of praise and blame, as well as the legal practice of reward and punishment, we admit several excusing conditions and mitigating factors. If these practices involve the ultimacy condition, then very little if any of excusing conditions and mitigating factors are accepted. This is because we would be very reluctant to accept excusing conditions or mitigating factors if we accept that we are ultimately responsible not only for all our choices and actions, but for our wills, character traits, and motives as well. The ultimacy condition ends up making us morally responsible for more than we actually deserve.

**Moral Responsibility without Ultimacy Condition**

In their own ways, Nagel and Strawson attempted to show us that retrospective moral responsibility is impossible, no matter if causal determinism is true or false. While Nagel points out that moral luck robs us of the control we think we have, Strawson points out that we cannot be truly responsible for the way we are. And since we have to be truly responsible for the way we are before we can be truly responsible for our choices and actions, we are not truly responsible for our choices and actions as well. Frankfurt proposes that we are morally

responsible for the choices and actions we really wanted to do and did, and not for whether we could have done otherwise. Peter Strawson suggested that our ordinary interpersonal attitudes can only be justified by considerations internal to our general framework of human practices and not by theoretical considerations (like the thesis of causal determinism) external to this general framework, and thus concludes that causal determinism, whether true or false, is irrelevant to our practices moral responsibility ascription. Robert Kane argues that we can only be ultimately responsible for our choices and actions if we can be ultimately responsible for our wills, character traits, and motives. And to be ultimately responsible for our wills, character, and motives, we must possess alternative possibilities in our past histories to form and shape our present wills, character traits, and motives.

I agree with Nagel that the control we have over our choices and actions are constrained by moral luck but disagree that moral luck would completely reduce our choices and action to mere occurrences. I agree with Galen Strawson that 'true' or 'ultimate' moral responsibility is impossible given the basic argument but disagree that the basic argument rules out partial moral responsibility. I agree with Frankfurt that we are morally responsible for the choices and actions we really wanted to do and did, and not for whether we could have done otherwise, but add that this view is compatible with a less robust sense of moral responsibility, and not the absolute or ultimate conception of it. I agree with Peter Strawson that our participant reactive attitudes needs to and can be defended, but disagree that the justification for our participant reactive attitudes must be internal to our general framework of human life and that external 'rational' justifications are irrelevant. I agree with Kane that in order for us to be ultimately responsible for our choices and actions, we have to be ultimately responsible for our wills, character traits, and motives; but disagree that ultimate responsibility for our wills, character traits, and motives, and for our choices and actions in

turn, is possible, whether causal determinism is true or false (that is, I agree with Galen Strawson's assessment of UR).

My own view again is that the agent's possessing a certain CPM and CPR, who engages in a certain process of practical reasoning (exercising his CPR), at a certain time or duration of time (a type of event), is treated as a co-influent or co-determinant of other; and is morally responsible only insofar as they are co-influents and co-determinants. This implies that agents can be morally praised or blamed, or legally rewarded and punished, only insofar as they are co-influents and co-determinants. For this reason, choices and actions are not entirely up to the agent, and so the agent cannot be morally responsible (and hence be praised, blamed, rewarded, or punished) for them in the absolute or ultimate sense. This view is committed to neither causal determinism nor causal indeterminism but remains consistent with both. If causal determinism is true, then there are sufficient antecedent conditions for the agent's CPM, CPR, and process of practical reasoning. The absence of alternative possibilities can only ground a less robust (non-ultimate) sense of moral responsibility. As argued under my discussion of Frankfurt above, we seem to hold people morally responsible for their choices and actions when they follow from their CPM, CPR, and process of practical reasoning, even if these factors have sufficient causal conditions, and not because they could have done otherwise, which implies that their choices and actions may not follow from their CPM, CPR, and process of practical reasoning. And if causal indeterminism is true, then there are insufficient antecedent conditions for the agent's CPM, CPR, and process of practical reasoning. The presence of alternative possibilities can only ground a less robust (non-ultimate) sense of moral responsibility. But even if the agent's CPM, CPR, and process of practical reasoning have no sufficient causal conditions, it may still be true that these factors contribute towards their choices and actions, and they can hence be held morally

responsible for them in a less robust (non-ultimate) way. Either way, UR, or a robust sense of moral responsibility is impossible. However, as long as our CPM, CPR, and process of practical reasoning, in the absence of coercion, compulsion, and constrain, are antecedent conditions (whether sufficient or insufficient) of our choices and actions, a less robust sense of moral responsibility (R without U) remains possible. And I believe that responsibility without ultimacy is not only good enough, but also more plausible for our moral practices of praise and blame as well as legal practices of reward and punishment. I have argued for this in my discussion of Robert Kane above that UR ends up making us morally responsible for more than we actually deserve.

Even if we accept that a less robust sense of responsibility without ultimacy is plausible, and accept that it is consistent with both causal determinism and causal indeterminism, we may still want to question the relationship between moral responsibility and causal responsibility. As discussed above, I reject Peter Strawson's view that metaphysical theses (whether causal determinism or causal indeterminism) are irrelevant to the justification of our participant reactive attitudes (or ascriptions of moral responsibility), and I accept the contrary view that our metaphysical theses (whether causal determinism or causal indeterminism) are relevant to the justification of our participant reactive attitudes (or ascriptions of moral responsibility). Without intending to argue for it here, I would like to propose that the relationship between moral responsibility and causal responsibility is this: causal responsibility is what grounds our moral practices of praise and blames as well as legal practices of reward and punishment. That is, our moral practices of praise and blame as well as our legal practices of reward and punishment can never be justified without causal responsibility.

**References:**


Eshlemen, Andrew (2009) 'Moral Responsibility', *Stanford Encyclopedia of Philosophy* (SEP) http://plato.stanford.edu/entries/moral-responsibility

Frankfurt, Harry (2009) 'Alternate Possibilities and Moral Responsibility' in *Free Will 2nd Edition* ed. Gary Watson (Oxford University Press)

Kane, Robert (2009) 'Free Will: New Foundations for an Ancient Problem' in *Free Will* ed. Derk Pereboom (Hackett)

Kane, Robert (1996) *The Significance of Free Will* (Oxford University Press)

Kane, Robert (1995) 'Two Kinds of Incompatibilism' in *Agents, Causes, and Events* ed. Timothy O'Connor (Oxford University Press)

Nagel, Thomas (1979) *Mortal Questions* (Cambridge University Press)

Strawson, Galen (2009a) 'The Impossibility of Moral Responsibility' in *Free Will 2nd Edition* ed. Gary Watson (Oxford University Press)

Strawson, Galen (2009b) 'The Impossibility of Ultimate Moral Responsibility' in *Free Will* ed. Derk Pereboom (Hackett)

Strawson, Peter (2009) 'Freedom and Resentment' in *Free Will 2nd Edition* ed. Gary Watson (Oxford University Press)

Widerker, David (2002) 'Responsibility and Frankfurt-type Examples' in *The Oxford Handbook of Free Will* ed. Robert Kane (Oxford University Press)

Williams, Garrath (2009) 'Responsibility', *Internet Encyclopedia of Philosophy* (IEP) http://www.iep.utm.edu/responsi

# CHAPTER 5: CAUSAL RESPONSIBILITY GROUNDS MORAL RESPONSIBILTY

## Introduction

In this essay, I aim to argue for the thesis that retrospective causal responsibility grounds moral responsibility that applies to cases of intended action, unintended action, and intended omissions, but not unintended omissions. First, I outline an account of causal responsibility, retrospective moral responsibility, and their relation. Second, I discuss and respond to J.J.C. Smart's utilitarian view that the ascription of retrospective moral responsibility is given up for the more useful ascription of prospective moral responsibility, and P.F. Strawson's normative view (as I attribute to him) that causal responsibility ascription is irrelevant to retrospective moral responsibility ascription. Third, I discuss and respond to the libertarian, hard incompatiblist, as well as negative causation objections against the retrospective causal responsibility grounds moral responsibility thesis. Fourth, I highlight the limitations of the retrospective causal responsibility grounds moral responsibility thesis but maintain that this does not reduce its plausibility.

## An Account of Causal Responsibility, Moral Responsibility, and Their Relation

Causal responsibility is an ontological relation between at least two events. For event X to be causally responsible for event Y, event X has to be both the efficient and proximate cause of event Y. I take efficient cause to mean the following: X is the efficient cause of Y, if the occurrence of X makes a difference to the occurrence of Y (difference-making account), or if X contributes towards the occurrence of Y (contributory account), or if X had not occurred then Y would not have occurred (counterfactual-supporting account), or if X is at least an

'Insufficient but Non-redundant' part of an 'Unnecessary but Sufficient' condition of Y (Mackie's INUS account). And I take proximate cause to mean the following: X is the proximate cause of Y, if X is the efficient cause of Y and if the occurrence of X immediately precedes Y, or alternatively, if X is the efficient cause of Y and if the occurrence of Y immediately succeeds X. The proximity requirement is necessary to exclude remote causes and background conditions. And the qualifier 'immediately' applied to the terms 'preceding' or 'succeeding' is rough and it need not specify any temporal limit. Causal responsibility, defined in terms of efficient and proximate causes, is then a broad and minimal notion of responsibility that is ascribable to both animate and inanimate objects exemplifying certain properties at some proximately earlier times, insofar as they 'make a difference to', 'counterfactually support', 'contribute toward', or 'are at least INUS conditions of' other animate or inanimate objects exemplifying certain properties at some proximately later times.

There are two distinct but related parts to causal responsibility when it is applied to persons. First, it describes the ontological relation between the bodily movements of persons at some earlier times (or durations of time) and other events at some later times (or durations of time). Thus applied, persons exemplifying certain bodily movements at some proximately earlier times 'make a difference', 'counterfactually support', 'contribute toward', or are at least 'INUS conditions of' other animate or inanimate objects exemplifying certain properties at some proximately later times. Bodily movements are construed broadly to include neural changes even when the body remains still. The events caused by the bodily movements of persons are the consequences or effects of actions. Let's call this the Action to Consequence Condition, or A→C Condition for causal responsibility ascription. Second, it describes the ontological relation between the intentional states of persons at some proximately earlier times and their bodily movements at some proximately later times. Thus applied, persons

exemplifying certain intentional states at some proximately earlier times 'make a difference', 'counterfactually support', 'contribute toward', or are at least 'INUS conditions of' themselves exemplifying certain bodily movements at some proximately later times. This is known as intentional causation. Intentional states are construed broadly to include desires and beliefs (before deliberation), as well as intentions (after deliberation). Let's call this the Person to Action Condition or P→A Condition for causal responsibility ascription.

If only the A→C Condition but not the P→A Condition part is taken into account, the event involving the bodily movements of persons may qualify as actions, but not as intentional actions. Only when both the A→C and P→A Conditions are taken into account, the event involving the bodily movements of persons caused by the intentional states qualify as intentional or volitional actions. To fulfil the P→A Condition for causal responsibility, a person must not only be the efficient and proximate causes, he must also be the intentional cause. I shall remain neutral between the relation of intentional states and neural states, be it dualism, identity, supervenience, or realisation. To be an intentional cause, a person needs to possess the capacity for practical reasoning. This is the capacity to deliberate and evaluate among desires and beliefs that eventually lead to intention and action (which usually involves bodily movement). Suppose B is an end, goal, purpose, or reason. A person desires B, believes that an action A leads to (or increases the probability of achieving) B, and after deliberating and evaluating among other existing desires and beliefs, intends to do A in order to B, and hence does A. Reasons are the contents of the person's intentional states. Intentional states like desires, beliefs, and intentions refer to the affective, cognitive, and conative attitudes respectively.

Moral responsibility is the normative relationship between 'a person X who performed an action Y with consequence Z' and 'others' attitudes (usually praise and blame) toward person

X for performing action Y with consequence Z'. It can also mean the normative relationship between 'a person X who performed an action Y with consequence Z' and 'person X's attitudes (usually praise and blame) towards himself/herself for performing action Y with consequence Z'. Action Y can be intended or unintended. Intended action involves the relevant bodily movement and the relevant intentional states with reasons as content. Unintended action involves the relevant bodily movement but not the relevant intentional states with reasons as content. Omission ~ Y can be intended or unintended too. Intended omission does not involve the relevant bodily movement but does involve the relevant intentional states with reasons as content. Unintended omission does not involve the relevant bodily movement and does not involve the relevant intentional states with reasons as content.

There are differing degrees of moral responsibility related to (1) whether or not an action and a consequence were intended, (2) whether or not the intended consequence occurred despite the action performed, or even (3) whether or not it is an action or an omission that leads to the intended consequence. On (1), if person X did not intend to perform action Y or did not bring about consequence Z, then person X is less morally responsible than if person X did intend to perform action Y or did bring about consequence Z. On (2), consequence Z may or may not follow from action Y, for there may be actions without consequences. Arguably, person X is less morally responsible if he did intend both action Y and consequence Z, but consequence Z did not occur (Thomson 1987). This point is controversial but I shall not discuss it here. On (3), some philosophers argue that omissions cannot be causes. I am going to contend that intended omissions, but not unintended ones, can be causes later. Putting this contention aside for now, it can still be arguably claimed that person X is less morally responsible if he omitted to perform action Y that resulted in intended consequence Z, rather than if he performed action Y that resulted in intended consequence Z.

A person is morally responsible for an action and its consequence when he is either praiseworthy or blameworthy for that action and its consequence. Related to the notion of praiseworthiness or blameworthiness is the notion of desert, where 'worthiness' roughly means 'deserving'. So, praiseworthiness and blameworthiness mean deserving of praise or blame. Analogous to causal responsibility, there are two distinct but related parts to moral responsibility when it is applied to persons. First, if person X is morally responsible for performing an action Y, then person X is either praiseworthy or blameworthy for performing action Y. This is the P→A Condition for moral responsibility ascription. Second, if person X performed action Y that brought about consequence Z, then person X is either praiseworthy or blameworthy for having brought about consequence Z. This is the A→C Condition for moral responsibility ascription.

My main contention is that retrospective causal responsibility grounds moral responsibility in cases of intended action, unintended action, and intended omissions. The grounding relation refers roughly to a justifying, rationalising, or warranting relation. There are two parts to the relation between causal responsibility and retrospective moral responsibility. The first part of the relation states that in cases of morally relevant actions, person X is blameworthy or praiseworthy for performing an action Y only if person X exemplifies the relevant intentional states at some proximately earlier time that 'make a difference to', 'counterfactually support', 'contribute toward', or 'are INUS conditions of' the person performing action Y at some proximately later time. Put more simply, a person X is morally responsible for performing action Y only if person X exemplifies the relevant intentional states that are causally responsible for the performance of action Y. That is, a person fulfils the P→A Condition for moral responsibility ascription only if the person fulfils the P→A Condition for causal responsibility ascription. The second part of the relation states that person X is either

blameworthy or praiseworthy for the occurrence of consequence Z, if the performance of action Y at some proximately earlier time 'makes a difference to', 'counterfactually supports', 'contributes toward', or 'is an INUS condition of' the occurrence of consequence Z at some proximately later time. That is, a person fulfils the A→C Condition for moral responsibility ascription only if the person fulfils the A→C Condition for causal responsibility ascription.

However, the P→A and A→C conditions are not sufficient to serve as objective conditions for moral responsibility ascription. For most compatibilists, a crucial condition is that action Y has to be caused in the right way by person X. I take 'caused in the right way by person X' to mean 'caused by person X when his capacity for practical reasoning is functioning properly'. That is, person X's capacity to deliberate and evaluate among desires and beliefs that eventually lead to intention and action is functioning properly, or person X's cognitive, affective, and conative capacities are functioning properly in such as way that they are responsive or sensitive to reasons. And I assume that when person X's capacity for practical reasoning is functioning properly; it is possible for him to foresee or predict that action Y leads to (or increases the probability of achieving) consequence Z. There are various impediments to the proper functioning of the person's capacity for practical reasoning. The first two pertains to psychological factors internal to the person, which refer to immaturities and abnormalities. Demarcating between clear and borderline cases of immaturities and abnormalities is not without problems, but it seems plausible to hold that a person is less morally responsible if he falls under the borderline cases, and that a person is not morally responsible if he falls under clear cases. The next two pertains to psychological factors external to the person, which refer to coercion and manipulation by others. Likewise, demarcating between clear and borderline cases of coercion and manipulation is not without

problems, but it seems plausible to hold that a person is less morally responsible if he falls under the borderline cases, and that a person is not morally responsible if he falls under clear cases. The notion of manipulation that I am considering here is local, that is, it refers to a particular or specific social condition the person is in. The other notion of manipulation is global, that is, it refers to universal or general condition of being causally determined by forces beyond the person's control. I will address this stronger notion later. Putting this stronger notion aside for now, let's call the condition where a person's capacity for practical reasoning is functioning properly the No-Impediment Condition or Negligible-Impediment Condition, or N-I Condition.

So far, I have discussed the three objective conditions of causal responsibility ascription that are necessary to justify moral responsibility ascription in cases of intended action, unintended action, and intended omission. They are the: (1) P→A Condition, (2) A→C Condition, and (3) N-I Condition. While I do claim that (1) to (3) provide the objective (external) causal conditions for the ascription of moral responsibility, I do not claim that the conjunction of all three objective conditions are necessary or that they are jointly sufficient. Here is why. An intended action that caused an intended consequence satisfies all conditions (1) to (3). For example, a driver intends to kill a pedestrian by running him over with his car, hit him and killed him. An intended action that did not cause an intended consequence satisfies conditions (1) and (3) but not (2). For example, a driver intends to kill a pedestrian by running him over with his car, but missed him and hence failed to kill him. An unintended action that caused an unintended consequence satisfies conditions (2) and probably (3) but not (1). For example, a driver failed to see a pedestrian crossing the road in front of him and to stop his car in time, and accidentally hit him and killed him. An unintended action that did not cause any consequence satisfies probably only condition (3). For example, a driver failed

to see a pedestrian crossing the road in front of him and stop his car in time, but narrowly missed him and hence did not kill him. So, all cases except unintended action that did not cause any consequence justify moral responsibility ascription to different degrees. And it seems that the conjunction of at least conditions (1) and (3) or conditions (2) and (3) justifies moral responsibility ascription. It seems unclear whether the conjunction of conditions (1) and (2) but not (3) justifies moral responsibility ascription. That is, whether it is justifiable to ascribe moral responsibility in cases where person X is the efficient, proximate, and intentional cause of action Y but his capacity for practical reasoning is not functioning properly. It depends. If condition (3) is not satisfied because of clear cases of immaturity, abnormality, coercion, and manipulation, then I am inclined to answer in the negative because we do not usually ascribe moral responsibility to persons who lacked adequate control over their capacity for practical reasoning. However, if condition (3) is not satisfied because of borderline cases of immaturity, abnormality, coercion, and manipulation, or because of clear cases of impulsive or reckless behaviour, then I am inclined to answer in the affirmative because we do usually ascribe moral responsibility to persons who have adequate control over their capacity of practical reasoning. Satisfaction of Condition (3) in cases of global manipulation will be discussed further below.

In cases of omissions, an intended omission that caused an intended consequence satisfies conditions (1) and probably (3) but not (2). For example, a lone bystander noticed a drowning swimmer in the pool, is able to help but is unwilling to do so because he desires the swimmer's death, and the swimmer drowned. Similarly, an intended omission that did not cause an intended consequence satisfies conditions (1) and probably (3) but not (2). For example, a lone bystander noticed a drowning swimmer in the pool, is able to help but is unwilling to do so because he desires the swimmer's death, but the swimmer survived

somehow. An unintended omission that is connected to an unintended consequence satisfies probably only condition (3). For example, a bystander did not notice the drowning swimmer (perhaps being distracted for some reason) and the swimmer drowned, but he would be able and willing to help if he noticed the drowning swimmer. Similarly, an unintended omission that is not connected to any consequence satisfies probably only condition (3). So, it seems that all cases except unintended omission justifies moral responsibility ascription to different degrees. And yet, there seems to be cases where an unintended omission that is connected to an unintended consequence justify moral responsibility ascription, even when there seems to be no ascription of causal responsibility, like cases of negligence. Cases of negligence, construed as forms of unintended omissions, are perhaps the only exception to the causal responsibility grounds moral responsibility thesis, unless they can be construed as cases of intended omission. I shall discuss this further below. And generally, a person who intended an omission is less responsible than a person who intended an action. This is because an action requires more effort and is more likely to cause a consequence than an omission. In his paper 'Saving Life and Taking Life', Richard Trammell argues for a distinction between negative and positive duties (actions and omissions) as follows: "The negative duty not to kill can be fully discharged, whereas the duty to save cannot. Failure to meet the duty of not killing cuts off any possibility of realizing the good connected with the life in question, whereas failure to save leaves open the option for someone else to save. Finally, a person is not necessarily responsible for someone else's needing to be saved; but he is responsible for the life of anyone he kills" (Trammell 1975, pp.136). But it is not my aim to argue for this claim here. My aim is to show that subjective (internal) conditions – our shared moral values and norms – alone are never sufficient for most cases, not to list all the conditions required.

I contend that the normative notion of moral responsibility can have its basis on the ontological notion of causal responsibility, which provides objective (external) conditions for the ascription of moral responsibility based on causal relations between events, on top of subjective (internal) conditions based on certain shared moral values and norms. Objective (external) conditions are independent of any existing moral (normative) practices or standards while subjective (internal) conditions are dependent on some existing moral (normative) practices or standards. It is important to note that my contention takes no stance between causal determinism and causal indeterminism. All the causal theories adopted here, the difference-making account, counterfactual account, contributory account, or INUS account, are compatible with both causal determinism and causal indeterminism. My contention is that the ascription of causal responsibility grounds moral responsibility, be it deterministic or indeterministic. If it is deterministic, then there are sufficient antecedent causal conditions for the occurrence of events including actions and their consequences. If it is indeterministic, there are insufficient antecedent causal conditions for the occurrence of events including actions and their consequences. In cases of causal responsibility involving persons, my account assumes a version of intentional causation. That is, a person exemplifying the relevant intentional states with relevant reasons as their content is causally responsible for his actions or omissions. However, my contention is highly controversial even among some compatibilists, who claim that there are other ways to understand moral responsibility that do not require causal responsibility.

**Smart and Strawson on Moral and Causal Responsibility**

There are compatibilists who have argued against the need for retrospective moral responsibility and the need for objective (external) conditions for the ascription of moral

responsibility. I will briefly discuss and eventually put aside two such compatibilist accounts of justifying the ascription of moral responsibility before proceeding further with my own causal integrationist account. Both accounts attempt to side-step the compatibility question between causal determinism (it may well be causation in general) and moral responsibility. That is, they imply that moral responsibility does not require causal responsibility. The first account is proposed by J.J.C. Smart in his article 'Free Will, Praise, and Blame'. And the second account is proposed by P.F. Strawson in his article 'Freedom and Resentment'.

Smart affirms that moral responsibility should only be ascribed to those candidates with a capacity to modify their behaviour once moral responsibility is ascribed to them. To illustrate this point, he first draws a comparison between a stupid boy and a lazy boy (Smart 2009, pp.68). If the stupid boy does not do his homework, it would not be appropriate for the schoolmaster to hold him morally responsible because doing so will not make him more intelligent. And the schoolmaster may well conclude that the stupid is not blameworthy because he could not have done his homework even if he had tried. Whether the stupid boy does his homework is largely influenced by hereditary and environmental factors and there is nothing much the schoolmaster can do to change it. Note that this point is contentious because it is true only if causal determinism by external factors is assumed. If causal determinism by external factors is not assumed, the stupid boy can still be praiseworthy for trying hard and blameworthy for not trying hard but I will not pursue it here. However, if the lazy boy does not do his homework, it would be appropriate to hold him responsible because doing so may make him more hardworking in the future. And the schoolmaster may well conclude that the lazy boy is blameworthy because he could have done his homework if he had tried. Whether the stupid boy does his homework is largely influenced by hereditary and environmental factors and the schoolmaster constitutes part of the environmental factors.

This leads Smart to further affirm that ascription of moral responsibility has a clear pragmatic justification in regulating prospective behaviour that is consistent with causal determinism. And it may be true that the ascription of prospective moral responsibility requires causal determinism, since the pragmatic and regulating function of prospective moral responsibility ascription will be ineffective if causal determinism is false. Smart has little to say about the ascription of retrospective moral responsibility and whether it is consistent with causal determinism. However, his view concerning the justification of responsibility ascription does suggest that the ascription of retrospective moral responsibility serves no useful purpose and hence the question about the relationship between retrospective moral responsibility and causal determinism (or any causal theories in general) is being side-stepped.

Strawson defines his position on the ascription of moral responsibility by responding to the deficiencies of what he calls the optimist's and the pessimist's views. The optimist's view is one like Smart's account where the practices of ascription of moral responsibility are justified by its efficacy in regulating behaviour in socially desirable ways. In other words, such practices are treated as nothing more than instruments of individual treatment and social control. The optimist claims that these practices require us to inhibit or suspend our participant reactive attitudes in general and adopt the objective attitude with respect to ascriptions of moral responsibility. And they are consistent with the truth of causal determinism. The pessimist's view is like the libertarian's account in general. It agrees that our participant reactive attitudes should be inhibited or suspended, or that we should adopt the objective attitude, when responding specifically to abnormalities (people with psychological disorders) or immaturities (children), but contends that we cannot do so when we relate to others in general, when abnormalities or immaturities are absent. As a whole, the pessimist does not challenge the usefulness of prospective moral responsibility ascription

proposed by the optimists, but insist that it is neither a sufficient nor right basis for our participant reactive attitudes as we understand them. Our practices of moral responsibility ascription are not justified by their efficacy in regulating behaviour in socially desirable ways. Rather, they are justified because those who are morally responsible are guilty of wrongdoing and they deserve to be blamed or punished. And this vital part of our moral practices is ignored by the optimists. The pessimist claims that our practices are grounded retrospectively rather than prospectively and that retrospective moral responsibility is inconsistent with causal determinism.

Strawson remarks that both the optimist and pessimist misconstrue the facts in different ways by over-intellectualising them. Questions about the justification of our practices of moral responsibility ascription are internal to the structure of our participant reactive attitudes or relate to modifications internal to it. This existence of this general framework of human attitudes itself is something we are given with the existence of human society. It neither calls for nor permits an external 'rational' justification. And both the optimist and pessimist are unable to accept this. The optimist is right in seeking to find an adequate basis for social practices in calculated consequences, but wrong in losing sight of the fact that these social practices are expressions of human attitudes and not merely calculative devices employed for regulative purposes. The pessimist is right acknowledging the significance of this fact, but wrong in not accepting that inclusion of the human attitudes alone is sufficient to fill the gap in the optimist's account. This leads the pessimist to invent what Strawson calls the 'obscure and panicky metaphysics of libertarianism' (Strawson 2009, pp.93) in order to justify moral responsibility ascription. For Strawson then, the optimist's view is the right one if it is sufficiently modified by accounting for human attitudes. The inclusion of the pessimist's libertarian metaphysics is entirely unwarranted because our practices of moral responsibility

ascription are subject to internal justification, not external 'rational' justification (like causal determinism). For him, the ascription of retrospective moral responsibility is too embedded in us to be dislodged by our theoretical convictions (like causal determinism). And by holding that external 'rational' justifications or theoretical convictions are irrelevant to the ascription of retrospective moral responsibility, the question about the relationship between retrospective moral responsibility and causal determinism (or any causal theses in general) is being side-stepped.

Like Strawson, I agree with Smart that ascription of prospective moral responsibility is useful to have. And like Strawson but unlike Smart, I agree that ascription of prospective moral responsibility does not explain or justify our participant reactive attitudes, and even that the ascription of retrospective moral responsibility may be too embedded in us to be dislodged by our theoretical convictions. But contra Strawson, I hold not only the weaker claim that external 'rational' justifications or theoretical convictions are relevant to the ascription of retrospective moral responsibility, but the stronger claim that external 'rational' justifications or theoretical convictions ground the ascription of retrospective moral responsibility as well. Since what is in contention is the retrospective aspect (and not prospective aspect) of moral responsibility, I shall mean retrospective moral responsibility whenever I use term moral responsibility from now on. Strawson's normative account of moral responsibility holds that agents are morally responsible insofar as it is fair to ascribe moral responsibility or apply participant reactive attitudes to them. And whether it is fair to ascribe moral responsibility or apply participant reactive attitudes to agents is internally determined by prevailing normative practices, and not externally by any naturalistic considerations. Although I concede that prevailing normative practices usually influence whether it is fair to hold agents morally responsible, I dispute that our prevailing normative practices alone can justify ascriptions of

moral responsibility without naturalistic considerations. And I disagree with the claim that adopting the objective attitude (unless in cases of abnormalities or immaturities) requires us to inhibit or suspend our participant reactive attitudes or ascriptions of moral responsibility in general, whether causal determinism or causal indeterminism is true.

There are two reasons why I disagree with Strawson's account. The first is Strawson's claim that we cannot adopt an objective attitude under normal and mature conditions. There is no good reason to adopt such a narrow sense of objective attitude that excludes normal and mature conditions, for it is possible to adopt broader sense of objective attitude that encompasses a person's causal responsibility for his actions and their consequences. And I believe that this broader sense of objective attitude forms a fairer basis for ascribing moral responsibility than basing our reactive attitudes solely on the prevailing normative practices. Such an objective attitude provides a fairer basis because we ascribe moral responsibility to a person only insofar as he is causally responsible (whether causal determinism or causal indeterminism is true) for his actions and their consequences, for there seems to be no justification for ascribing moral responsibility to a person if he makes no difference to, does not counterfactually support, does not contribute to, or is not an INUS condition of, his action and its consequences. The second is Strawson's claim that that our reactive attitudes may be too embedded in us to be dislodged by theoretical convictions. This may be an accurate description of our psychological inclinations and tendencies pertaining to reactive attitudes, but it need not rule out the possibility that our theoretical convictions are relevant to the justification of our reactive attitudes. It is not implausible to ask whether our reactive attitudes should be revised if not dislodged given plausible theoretical convictions. And it is not implausible to ask whether a person's causal responsibility for his actions and their consequences should be considered before we apply our reactive attitudes toward him. If the

answers to these questions are yes, then it is not implausible to adopt an objective attitude in the broader sense as mentioned above and affirm that external 'rational' justification are relevant to, or even required for, the ascription of moral responsibility. Moral responsibility ascription may not need 'obscure or panicky metaphysics', but this does nothing to show that moral responsibility ascription does not need metaphysics at all.

**Libertarians on the Causal Responsibility grounds Moral Responsibility Thesis**

Some libertarians assert that objective (external) conditions, like causal responsibility, are not required for moral responsibility ascription. They are known as simple indeterminists or non-causalists, or non-causal libertarians. They believe that some actions are free, free actions are uncaused, and uncaused actions are done for reasons. Perhaps moral responsibility ascription requires only free, uncaused actions done for reasons because they fall outside the causal order. Hence, moral responsibility does not entail causal responsibility. The problem with this view is that action, whether free or not, involves bodily movement. And if action is done for reasons, then it involves voluntary bodily movement. The claim that voluntary bodily movement is uncaused, or that 'nothing makes a difference to', 'nothing counterfactually supports', 'nothing contributes toward', 'nothing is an INUS condition of' voluntary bodily movement, seems unintelligible. Non-causal libertarians can reply that reasons are what explain actions and render them intelligible. But since reasons do not figures as causes, they 'do not make a difference to', 'do not counterfactually support', 'do not contribute towards', or 'are not INUS conditions of' voluntary bodily movements. Ascribing moral responsibility or attitudes of praise and blame to persons for uncaused (not even by their own relevant intentional states with relevant reasons as their content) voluntary bodily movements seems unintelligible and unjustified. This seems to be an undesirable outcome for libertarians.

Other libertarians assert that objective conditions (1) to (3) are not necessary or sufficient for the kind of causal responsibility that is required for the ascription of moral responsibility. They are divided into two camps: the event-causal libertarians and the agent-causal libertarians. Both the event-causal and agent-causal libertarians seem to be committed to the ultimate conception of moral responsibility. And I assume that non-causal libertarians are committed to it as well since they affirm that free actions fall outside the causal order. Ultimate moral responsibility refers to the complete, full, or sole, attribution of moral responsibility, or to apply attitudes of praise and blame completely, fully, or solely to persons for performing free actions that result in consequences. For both the event-causal and agent-causal libertarians, this is justified by ultimate causal responsibility involving persons. This means that persons are grounds, origins, or sources of free actions. Put in another way, free actions must be 'entirely up to us', and have their grounds, origins, or sources 'entirely in us'. Both camps are committed to the ultimate moral responsibility entails ultimate causal responsibility thesis, but they differ in their characterisation of ultimate causal responsibility.

For the event-causal libertarians, causal indeterminism is a necessary condition for ultimate moral responsibility ascription. That is, ultimate moral responsibility entails indeterministic causal responsibility. And this means that on top of objective conditions (1) to (3), person X must possess the ability to act otherwise (performing action non-Y instead of action Y) given identical process of practical reasoning, as well as identical intentional states and reasons. Let's call this objective condition (4). Event-causal libertarians can accept that person X is causally determined to perform action Y only if person X's process of practical reasoning and relevant intentional states with the relevant reasons as content leading to action Y is not causally determined. The first problem with (4) is that it is not sufficient for ultimate causal responsibility. Ultimate moral responsibility entails ultimate causal responsibility, not just

indeterministic causal responsibility. If person X's process of practical reasoning, intentional states, and reasons only probabilistically or statistically determine action Y, then the remaining causal gap is filled with luck or random factors beyond the control of person X. Performance of action Y is then no longer entirely up to person X, as it does not have its ground, origin, or source entirely in person X. It follows then that person X does not bear ultimate causal responsibility for action Y and hence does not bear ultimate moral responsibility for action Y. At best, indeterministic causal responsibility can support non-ultimate (adequate or proximate) moral responsibility as long as person X's process of practical reasoning, intentional states, and reasons cause ('make a difference to', 'counterfactually support', 'contribute toward', or 'are an INUS condition of') action Y, even if it does not determine it. This brings us to the second problem with (4), which is that it is not necessary for non-ultimate (adequate or proximate) moral responsibility. For non-ultimate moral responsibility is consistent with causal determinism too, so far as person X's process of practical reasoning, as well as intention states with reasons as content, are co-determinants of action Y, with other co-determinants beyond the control of person X (just like luck or random factors beyond the control of person X under causal indeterminsm).

For the agent-causal libertarians, agent (as substance) causation is a sufficient condition for ultimate moral responsibility ascription. That is, ultimate moral responsibility entails agent (as substance) causal responsibility. On this account, person X (as substance) causes action Y by freely exercising his volitional-enabling active powers. The term 'freely' implies that nothing causes person X (as substance) to exercise his volitional-enabling powers. In short, person X (as substance) is a sufficient causal condition of action Y. Let's call this objective condition (5). If adopted, (5) displaces the causal work done by objective conditions (1) and (2). Objective condition (3) is rendered irrelevant because 'caused in the right way by person

X' now means 'caused by person X who freely exercises his volitional-enabling powers'. And objective condition (4) is too rendered irrelevant because the performance of action Y or non-Y rests entirely with person X (as substance). The strength of (5) is that it seems sufficient for ultimate causal responsibility, where performance of action Y is entirely up to person X, and does seem to have its ground, origin, or source entirely in person X. It is person X who considers his intentional states with reasons as content that causes his actions. Intentional states with reasons as content are not causes, only person X (as substance) is. And in virtue of this, it seems justified to attribute ultimate moral responsibility to person X for performing action Y. The weakness of (5) is unintelligibility, for what it is about person X that causes action Y remains mysterious and inexplicable. Agent-causal libertarians are committed to saying that nothing causes ('makes a difference to', 'counterfactually supports', 'contributes toward', or 'is an INUS condition of') person X to perform action Y. Of course, they can cite reasons to make action Y intelligible, but since they affirm that reasons do not figure as causes, then they 'do not make a difference to', 'do not counterfactually support', 'do not contribute towards', or 'are not INUS condition of' action Y. And since intentional states do not figure as causes either, this implies that person X may still perform action non-Y rather than action Y given identical processes of practical reasoning, intentional states and reasons. In response to this, they can treat reasons as indeterministic structural causes instead. But this does not help as no matter what objective probabilities, propensities, or tendencies reasons may confer towards the performance of action Y or action non-Y, person X (as substance) can still overrule reasons by freely exercising his volitional active powers. So, the question what it is about person X that causes action Y remains mysterious and inexplicable. And it seems that person X (as substance) is simply stipulated as an uncaused causer, an explanation that cannot itself be explained.

All libertarian positions discussed rely on some form of uncaused event or substance as a requirement for moral responsibility attribution. While it is possible for there being an uncaused event or substance, it seems neither rational nor justified to attribute moral responsibility based on it. As I have discussed above, ignoring the relevance of causal responsibility for moral responsibility has problems, and the same goes for asserting that objective conditions (4) or (5) are the kind of causal responsibility necessary or sufficient for moral responsibility ascription. The stronger thesis 'ultimate moral responsibility entails ultimate causal responsibility', as held by libertarians, is both naturalistically implausible and normatively unnecessary. The notion of ultimate causal responsibility is naturalistically implausible because it requires the notion of uncaused self-causation (causa sui), that is both logically and empirically implausible. Logically speaking, to cause the way our selves are (of which our character, personality, motivational structure follow) already presupposes that our selves are already existing in some way (having some character, personality, and motivational structure). As Thomas Nagel explains in his paper 'Freedom': "…in order to do anything we must already be something" (Nagel 2009, pp.237). Empirically speaking, we do not have complete or full control over the circumstances that affect the development of our character, personality, and motivational structure. A simple argument can now be given against the stronger thesis based on the impossibility of ultimate causal responsibility: ultimate moral responsibility entails ultimate causal responsibility, and since ultimate causal responsibility is impossible, ultimate moral responsibility is impossible. Here, some libertarians may want to give up the stronger thesis in favour of the weaker thesis 'causal responsibility grounds moral responsibility'. But if they do, then their account has no distinct advantage over the compatibilists, who can easily accommodate the weaker thesis. I hope to have shown that objective conditions (1) to (3) are sufficient for a weaker notion of causal responsibility, which is able to provide objective conditions to ground a weaker notion of moral

responsibility attribution. The stronger notion of moral responsibility is normatively unnecessary because there is no justification for attributing moral responsibility to a person for circumstances that affect the development of his character, personality, and motivational structure that are beyond his control. Attributing a weaker notion of moral responsibility by taking into account his capacity for practical reasoning, process of practical reasoning, and his relevant intentional states and reasons that cause his actions, seems adequate for our normative practices. Another simple argument can now be given against the ultimacy condition based on the non-necessity of ultimate moral responsibility: ultimate moral responsibility entails ultimate causal responsibility, and since ultimacy is not required for moral responsibility, ultimacy is not required for causal responsibility.

**Hard Incompatibilists on the Causal Responsibility grounds Moral Responsibility Thesis**

The hard incompatibilists, who hold that there is neither free will nor moral responsibility no matter whether causal determinism is true or causal indeterminism is true, formulated two powerful objections against the compatibilists which have significant bearings on the causal responsibility grounds moral responsibility thesis – the manipulation argument and the ultimacy argument. A version of the manipulation argument, as proposed by Derk Pereboom, is known as the Four-Case Argument (Pereboom 2009, pp.309-14). The first case features an agent who fulfils all the prominent compatibilist conditions for moral responsibility ascription, and yet owing to the fact that he is a victim of a direct manipulation (via timely remote control for example) of his process of practical reasoning, elicits the intuition that he is not morally responsible. The second case features an agent who fulfils all the prominent compatibilist conditions for moral responsibility ascription, and yet owing to the fact that he

is a victim of an indirect manipulation (via prior neural programming for example) of his process of practical reasoning, elicits the intuition that there is no salient difference between the first case and the second cases, and hence that he is not morally responsible. The third case features an agent who fulfils all the prominent compatibilist conditions for moral responsibility ascription, replaces the role that manipulation plays in the first and second cases with a very rigorous psychological or social conditioning of his home and community, and elicits the intuition that there is no salient difference between the first, second, and third cases, and hence he is not morally responsible. The fourth case features an agent who fulfils all the prominent compatibilist conditions for moral responsibility ascription, and causal determinism is true, and elicits the intuition that there is no salient difference between the first, second, third, and fourth cases, and hence that he is not morally responsible.

The main points of the Four-Case Argument is then to challenge the compatibilists to expose a salient difference between the fourth case and the other cases and explain why the agent is morally responsible in the fourth case but not in one or more of the other cases; and to elicit the intuition that there is no salient difference between the four cases and that the agent is not morally responsible in all the four cases despite fulfilling all the prominent compatibilist conditions for moral responsibility ascription. For this discussion, I will take the prominent compatibilist conditions for moral responsibility ascription to be conditions (1) to (3) as discussed above. I do not dispute that conditions (1) and (2) are satisfied in all the Four Cases, but whether condition (3) is also satisfied in all the Four Cases remains unclear. At first glance, it seems that condition (3) is not satisfied in the first case, but is satisfied progressively to different degrees from the second to fourth cases. I shall discuss this further below. Pereboom also holds that the agent is not morally responsible in a case where causal indeterminism is true because there is no salient difference between a manipulation case and

a case where causal indeterminism is true as well. In all the Four Cases, the agent has no control over the manipulative, deterministic, or indeterministic forces that condition his choice and action, and hence he bears no morally responsibility for them; or the agent is not the ultimate initiator, originator, or source of his choice and action, and hence he bears no moral responsibility for them.

According to Michael McKenna, two lines of reply to the Four-Case Argument are available to the compatibilists (McKenna 2008, pp.142). The Hard-Line Reply that McKenna prefers is to accept that there is no salient difference between the cases, and that the agent is morally responsible in all of the cases as long as the prominent compatibilist conditions for moral responsibility ascription are fulfilled. He formulated a four step Hard-Line Reply that consists of (1) rejecting all cases that do not fulfil the prominent compatibilist conditions, (2) improving on the manipulation cases to ensure that they fulfil the prominent compatibilist conditions, (3) focussing on the salient agential and moral properties that are relevant to moral responsibility ascription, and (4) making clear that non-manipulative causal factors and manipulation have the same effects but they undermine neither free will nor moral responsibility. In contrast, the Soft Line Reply claims that there are salient differences between the manipulative cases and the deterministic cases, and that the agent is not morally responsible in the manipulative cases where at least one prominent compatibilist condition for moral responsibility ascription is not fulfilled. There are a few plausible Soft-Line Replies to Pereboom's Four-Case Argument and McKenna's Four-Step Reply but I shall not discuss them here. It is worth noting that for McKenna, Soft Line Replies are bound to fail because better manipulative cases that are indistinguishable from deterministic cases can always be formulated to accommodate any fine-tuned compatibilist conditions of moral responsibility ascription. If Pereboom's and McKenna's arguments for the no-difference claim between

manipulative and deterministic cases are plausible, then there seem to be only two options available. Given the counter-intuitiveness of ascribing moral responsibility in the manipulative cases, either deny (with the hard incompatibilist and libertarians) that the prominent compatibilist conditions are adequate for moral responsibility ascription, or bite the bullet (with the hard-line compatibilists) and affirm that moral responsibility is ascribable to the manipulative cases as long as they fulfil the prominent compatibilist conditions which are adequate for moral responsibility ascription. Since there are strong arguments for the Hard-Line Reply despite its counter-intuitiveness, I would accept it only as a last resort or only if plausible Soft-Line Replies are not available. I shall discuss the bearing of the Manipulation Argument has on the causal responsibility grounds moral responsibility thesis and outline my own Soft-Line Reply to the Four-Case Argument below.

If the Four-Case argument succeeds in showing that there is no salient difference between manipulation cases and deterministic or indeterministic cases, and if an agent is not causally responsible for his choice and action in the manipulation cases, then he is not causally responsible for his choice and action in the deterministic or indeterministic cases. And if it is required that an agent is causally responsible for his choice and action before he is morally responsible for them, then he is not morally responsible for them. Put simply, if an agent is morally responsible, then he is causally responsible. The agent is not causally responsible. Therefore, he is not morally responsible.

One way to reject the Four-Case argument and the Hard-line Reply in order to defend the causal responsibility grounds moral responsibility thesis is to deny the no-difference claim. My rejection of the no-difference claim has two features. First, I focus on the intentional nature of the influents and determinants behind an intentional cause of an action, in particular whether and to what extent they impede the process of the agent's practical reasoning, and

whether they merely affect the content (or reasons that figure in the process) but not the process of the agent's practical reasoning. Second, I focus on the notion of control and moral responsibility ascription as a matter of degree rather than as a matter of either-or.

A manipulation case can be conceived as a case of a later intentional cause of action having an earlier intentional cause or causes. This applies to all of Pereboom's Four-Cases. In the first case, the earlier intentional cause (direct manipulation via timely remote control) can be conceived as a solely sufficient causal condition of the later intentional cause. Here, the earlier intentional cause completely impedes the process of practical reasoning behind the later intentional cause; objective condition (3) is unfulfilled; the manipulated agent possesses no control; and hence he bears no morally responsibility for his choice and action. In the second case, the earlier intentional cause (indirect manipulation via prior neural programming) can be conceived as a predominant but not a solely sufficient causal condition of the later intentional cause. Here, the earlier intentional cause severely impedes the process of practical reasoning of the later intentional cause (assuming that the earlier neural programming is not a solely sufficient causal condition of the later intentional cause); objective condition (3) is only minimally fulfilled; the manipulated agent possesses only minimal control; and hence he bears only minimal morally responsible for his choice and action. In the third case, the earlier intentional cause (very rigorous psychological or social conditioning of home and community) can be conceived as an influential but not a predominant or a solely sufficient causal condition of the later intentional cause. Here, the earlier intentional cause (assuming that the rigorous conditioning has no further effects similar to the first and second cases) significantly impedes the process of practical reasoning of the later intentional cause; objective condition (3) is only fulfilled in a limited way; the manipulated agent possesses limited control; and hence he bears only limited moral

responsible for his choice and action. In the fourth case, the earlier intentional cause (normal psychological and social conditioning of home and community) can be conceived as at most a negligible and not a significant, predominant, or sufficient causal condition of the later intentional cause. Here, the earlier intentional cause does not or only negligibly impedes the process of practical reasoning behind the later intentional cause; objective condition (3) is fulfilled; the causally determined (or causally indetermined) but not manipulated agent possesses adequate (but not complete or full) control; and hence is morally responsible (but not in the absolute or ultimate sense) for his choice and action. The agent in the fourth case is not morally responsible in the absolute or ultimate sense because even if there are no or negligible impediment (or manipulation) behind the process of the agent's practical reasoning, it is not possible for him to bear complete or full causal responsibility for all the influents and determinants behind his intentional causes. And given the causal responsibility grounds moral responsibility thesis, this means that it is not possible for the agent to bear absolute or ultimate moral responsibility for his action. However, it is my contention that adequate (or proximate) causal responsibility is sufficient to ground adequate (or proximate) moral responsibility, and that adequate (or proximate) moral responsibility is sufficient for our normative practices.

I have attempted to show that once we focus on the nature of the intentional influents and determinants behind an intentional cause of an action and on the notion of control and moral responsibility ascription as a matter of degree rather than as a matter of either-or, we are able see some salient differences between the Four Cases in terms of causal and moral responsibility ascriptions. And as long as some salient differences between the Four Cases can be identified, there is no conclusive reason to accept either the Manipulation Argument or the Hard Line Reply. In summary, the manipulation argument proceeds from the premises

that no one is morally responsible in manipulated cases, and that manipulated cases are analogous to causally determined (or causally indetermined) cases, to the conclusion that no one is morally responsible in causally determined (or causally indetermined) cases. My reply is that manipulated cases are not necessarily analogous to causally determined (or causally indetermined) cases and hence there are differences in the causal and moral responsibility ascriptions between them.

The manipulation argument is closely associated with, or even motivated by and based on the ultimacy argument, which argues from the premises that (P1) if causal determinism (or causal indeterminism) is true, then no one is the ultimate initiator, originator, or source of his choice and action, and that (P2) one bears moral responsibility for his choice and action only if he is the ultimate initiator, originator, or source of his choice and action, to the conclusion that (C1) if causal determinism (or causal indeterminism) is true, then no one bears moral responsibility for his choice and action. (C1) is a statement of incompatibility between causation (both deterministic and indeterministic) and moral responsibility. (P1) is plausible because no one can be the ultimate initiator, originator, or source of his choice and action if there are influents and determinants of his choice and action which are beyond his control, and there are always influents and determinants that affect the development of one's character, personality, and motivational structure (which in turn are the determinants and influents of his choice and action) which are beyond his control. (P2) is the questionable premise. I concede that one bears ultimate moral responsibility for his choice and action only if he is the ultimate initiator, originator, or source of his choice and action. And I affirm that one bears adequate moral responsibility for his choice and action only if he is the adequate initiator, originator, or source of his choice and action. But I deny that one bears adequate moral responsibility for his choice and action only if he is the ultimate initiator, originator, or

source of his choice and action. One does not have to be the ultimate initiator, originator, or source of his choice and action in order to bear adequate moral responsibility for them. Hence, I deny (C1) of the ultimacy argument by denying (P2). To summarise, ultimate moral responsibility requires ultimate sourcehood and adequate moral responsibility requires adequate sourcehood, but adequate moral responsibility does not require ultimate sourcehood. Causal determinism (or causal indeterminism) need not rule out adequate moral responsibility even if it rules out ultimate moral responsibility. Incompatibilists may insists that moral responsibility requires ultimacy but I have shown that ultimate moral responsibility requires ultimate causal responsibility, and since one cannot bear ultimate causal responsibility for his choice and action, he cannot bear ultimate moral responsibility for his choice and action. And if the incompatibilists concede that moral responsibility does not require ultimacy, then causal responsibility does not require ultimacy as well and this is compatible with causal determinism (or causal indeterminism).

In his book *Free Will*, Joseph Keim Campbell (Campbell 2011, pp.56-7 & 72) discusses an analogy between certain knowledge in epistemology and ultimate sourcehood in metaphysics. Claiming that knowledge requires certainty is analogous to claiming that sourcehood requires ultimacy. If knowledge requires certainty, then it can be easily shown that no one possesses knowledge. But no one cares about certainty because we accept that knowledge can be fallible and that fallible knowledge is more fitting to human capacities. Once we remove the need for certainty from knowledge, we can still affirm that we have fallible knowledge even if we deny that we have infallible knowledge. Analogously, if sourcehood requires ultimacy, then it can be easily shown that no one is the source of his choice and action. But no one cares about ultimacy because we accept that sourcehood can be adequate (without ultimacy) and that adequate sourcehood (without ultimacy) is more fitting to human capacities. In his

article ‘On Causation’, Van Rensselar Wilson (Wilson 1958, pp.231) points out the danger of supposing that an actual cause (or source) of an effect is somehow less a cause (or source) if it can be shown to have prior causes (or sources) of its own. This is analogous to supposing that I am somehow less the father of my children just because I in turn have a father, which sounds absurd. So, if I am the cause (or source) of my choice and action, I am no less a cause (or source) of my choice and action if I can be shown to have prior causes (or sources). Whatever moral responsibility I have for my choice and action, I have as their proximate (or adequate) cause (or source), and moral responsibility is not diminished if I can be shown to have prior causes (or sources). Wilson’s proposal can be made more plausible with one qualification – which is that moral responsibility can come without ultimacy. In his book *Causation and Responsibility*, Michael Moore (Moore 2009, pp.357-8) argues for the greater relevance of proximate cause and lesser relevance of remote cause to its effect by affirming the limited transitivity of causation as follows: if causation is scalar and that every event has multiple causes, then the more number of events there are between an effect and its putative cause, the less the putative cause contributes to the effect as compared to contributions of other causes. And at some point when the number of events between an effect and its putative cause are huge enough, then the contribution of the putative cause becomes so minimal that it would not count as a cause at all. That is, causation peters out so that even if A causes (is a source of) B, B causes (is a source of) C, and C causes…N, and yet for some N, A may well cause (be a source of) B or even C but not N. All three authors argue to the effect that the ultimacy or remoteness of cause (or source) either matters not or matters less as compared to the adequacy or proximity of cause (or source).

**Physical Connection and the Causal Responsibility grounds Moral Responsibility Thesis**

The problem of negative causation or the problem of omissions as causes in the context of action constitutes another powerful objection against the causal responsibility grounds moral responsibility thesis. According to this objection, actions can be causes but their negative, omissions, cannot be. And we sometimes attribute moral responsibility, or attitudes of praise and blame, to person X who omits to perform action Y that resulted in consequence Z. In cases of omissions, moral responsibility is attributed without causal responsibility. Hence, the causal responsibility grounds moral responsibility thesis is false. More specifically, the objection against negative causation is its incompatibility with the transference account of causation. On this account, there has to be a physical connection between a cause and its effect. A physical connection is usually conceived as a transfer of energy-momentum from a cause to its effect. Negative causes cannot exist because they are not physically connected to their effects, or because there is no energy-momentum transference from negative causes to their effects. In the context of action, negative causes refer to omissions. Action or bodily movement Y is physically connected to consequence Z and energy-momentum is transferred from action or voluntary bodily movement Y to consequence Z. The transference account of causation does not work for omissions because the A→C Condition is not satisfied in cases of omissions. Omission ~ Y is not physically connected to consequence Z, or energy-momentum is not transferred from omission ~ Y to consequence Z. Hence, omissions cannot be causes. I have two responses to this objection. The first is to reject the transference account of causation and accept omissions as causes on the other accounts of causation. And the second is to show that accepting omissions as causes is compatible with accounts of physical connection causation other than the transference account.

My first response is that omissions are compatible with the difference-making, contributory, counterfactual-supporting and INUS accounts of causation. Note these broader accounts can easily accommodate the narrower transference account but not vice versa. A→C Condition is not satisfied by omissions because there are no bodily movements involved in omissions. P→A Condition is satisfied by intended omissions. Person X → omission ~ Y means that person X exemplifying the relevant intentional states with relevant reasons as their content causes ('makes a difference to', 'counterfactually supports', 'contributes toward', or 'is an INUS condition of') omission ~ Y, or person X is the efficient, proximate, and intentional cause of omission ~ Y. Let's take a further look at the practical reasoning process behind intended omissions. Suppose an end, goal, purpose, or reason is represented by consequence Z. In cases of intended omissions with an intended consequence, person X desires consequence Z, believes that intended omission ~ Y leads to (or increases the probability of achieving) consequence Z, and after deliberating and evaluating among other existing desires and beliefs, intends omission ~ Y in order to achieve consequence Z, and hence permits omission ~ Y which in turn leads to consequence Z.

Negligent cases are problematic as the practical reasoning process behind negligent cases is not so clear cut. There are three ways of conceiving the practical reasoning process behind a negligence case. The first way is this: person X desires or intends not to exercise due care and diligence even when he has the belief (whether occurrent or dispositional) that omission ~ Y leads to (or increases the probability of achieving) consequence Z, and carelessly performs omission ~ Y which in turn leads to consequence Z. This way endorses the presence of relevant intentional states that motivates negligent omission. The second way is this: person X does not desire or intend to exercise due care and diligence even when he has the belief (whether occurrent or dispositional) that omission ~ Y leads to (or increases the

probability of achieving) consequence Z, and carelessly performs omission ~ Y which in turn leads to consequence Z. It is unclear whether this way endorses the presence or absence of relevant intentional states that motivates negligent omission. The third way is this: person X does not desire or intend to achieve consequence Z, but he may or may not have the belief (whether occurrent or dispositional) that omission ~ Y leads to (or increases the probability of achieving) consequence Z, and yet carelessly perform omission ~ Y which in turn leads to consequence Z. This way endorses the absence of relevant intentional states that motivates negligent omission. Assuming that the N-I Condition holds, that is, assuming that person X's capacity for practical reasoning is functioning normally, which implies that it is possible for him to foresee or predict that omission ~ Y leads to (or increases the probability of achieving) consequence Z. On the first way, it seems that cases of negligence are conceived as forms of intended omission, hence objective conditions (1) and (3) hold, and hence person X is causally responsible for omission ~ Y which leads to consequence Z. And the causal responsibility grounds moral responsibility thesis encompasses negligence cases. On the second way, it is unclear whether objective condition (1) holds even if objective condition (3) holds, hence it is unclear whether or not person X is causally responsible for omission ~ Y, and whether or not the causal responsibility grounds moral responsibility thesis encompasses negligence cases. On the third way, it seems that cases of negligence are conceived as forms of unintended omission, hence only objective condition (3) holds, and hence person X is not causally responsible for omission ~ Y which leads to consequence Z. And the causal responsibility grounds moral responsibility thesis does not encompass negligence cases. I shall remain neutral on which way of conceiving the practical reasoning process behind a negligence case is more plausible, but affirm that as long as negligence cases can be conceived as intended omissions, then the causal responsibility grounds moral responsibility thesis applies to them.

Two conclusions can be drawn from the claim that omissions can be accommodated by accounts of causation other than the transference account. The first is that these other accounts are mistaken and the second is that the transference account is limited. Here, I draw the second conclusion. While I admit that the transference account is useful at the more fundamental physical level, it construes the concept of cause too narrowly. And it has not been conclusively shown that the concept of cause cannot have broader applications as allowable by these other accounts. Moreover, these other accounts can explain why the transference of energy-momentum did not occur in cases of omission.

My second response states that negative causation, in the context of action like omissions, is compatible with accounts of physical connection causation other than the transference account. To causally connect an intentional state with a bodily state, some account of mind-body relation is required, whether an intentional state is identical with, supervenient on, or realised by, the neural state that causes it. I remain neutral towards these accounts of mind-body relation as I do not need to take sides to make my point. I just need to assume that at least one of them holds true. So, an intentional state can be identical with, supervenient on, or realised by, the neural state that causally connects with a bodily state. Usually, the bodily state that follows from an intentional/neural state is conceived as a voluntary bodily movement or an intentional action. From here, it is not hard to conceive the bodily state that follows from an intentional/neural state as a voluntary bodily restraint, in the form of an intended omission. If a physical consequence follows from a voluntary bodily restraint, in the form of an intended omission, which in turn follows from an intentional/neural state, then there seems to be an indirect physical causal connection between the intentional/neural state and the bodily state after all. And if this response is acceptable, then the intended omission as cause can be compatible with causal accounts that require physical connection but does not

require direct transference of energy-momentum from causes to effects. Assuming a properly functioning capacity for practical reasoning and further assuming that the psychological processes of practical reasoning are identical with, supervenient on, or realised by, the neural processes, objective conditions (1) and (3) are satisfied in cases of intended omission. Here, person X desires consequence Z, believes that omission ~Y leads to consequence Z, and intends omission ~ Y in order to achieve consequence Z (condition 1 fulfilled). It is also possible for him to foresee or predict that omission ~ Y leads to (or increases the probability of achieving) consequence Z (condition 3 fulfilled). Under the physical connection account, conditions (1) and (3) are identical with, supervenient on, or realised by, neural processes. Person X is then capable of voluntary action, conditions (1) and (3) are both fulfilled, and hence he is causally responsible for the omission ~ Y that leads to (or increases the probability of achieving) consequence Z. But as there is no physical connection between bodily states and their effects, objective condition (2) remains unfulfilled. Recall that my account only requires two out of three objective conditions to establish causal responsibility. So, for intended omissions under the physical connection (without transference) account, the conjunction of objective conditions (1) and (3) is sufficient to establish causal responsibility.

However, the objection can be pressed as follows. Even if causal responsibility is ascribable to cases of intended omission under the physical connection account of causation as described above, it is not ascribable to relevant cases of negligence. Yet, moral responsibility is ascribed to cases of negligence which leads to harmful consequences. Hence, the causal responsibility grounds moral responsibility thesis does not hold in cases of negligence on any account of physical connection causation. One response may be to look at the three ways of conceiving the practical reasoning process behind a negligence case as discussed above and assume that objective condition (3) holds. On the first way, there are relevant intentional

states identical with, supervenient on, or realised by, neural processes which are physically connected to the negligent omission, which in turn leads to its consequence. Here, objective condition (1) holds, voluntary bodily restraint is present, causal responsibility is ascribable, negligent cases are classifiable as cases of intended omission, and the causal responsibility grounds moral responsibility thesis is applicable. On the second way, it is unclear whether there are relevant intentional states identical with, supervenient on, or realised by, neural processes which are physically connected to the negligent omission, which in turn leads to its consequence. Here, it is unclear whether objective condition (1) holds, whether voluntary bodily restraint is present, whether causal responsibility is ascribable, whether negligent cases can be classified as cases of intended omission, and whether the causal responsibility grounds moral responsibility thesis is applicable. On the third way, there are no relevant intentional states identical with, supervenient on, or realised by, neural processes which are physically connected to the negligent omission, which in turn leads to its consequence. Here, objective condition (1) does not hold, voluntary bodily restraint is absent, causal responsibility is not ascribable, negligent cases can be classified as cases of unintended omission, and the causal responsibility grounds moral responsibility thesis is not applicable.

I have attempted to show that the physical connection accounts of causation other than the transference account are able to explain both the presence of energy-momentum transfer (in cases of intended and unintended action) as well as its absence (in cases of intended omission which may or may not include negligent cases), while the transference account is able to explain only the presence but not the absence of energy-momentum transfer. Given the broader applications of the physical connection accounts of causation other than the transference account, there is more reason to prefer them to the transference account, which has much narrower applications.

**Limitations of the Causal Responsibility grounds Moral Responsibility Thesis**

Despite having answered the above objections, my account of the causal responsibility grounds moral responsibility thesis has its limitations, but I shall argue that this in no way reduces its plausibility. The limitation is that causal responsibility, especially as stated in the form of objective conditions (1) to (3), is not sufficient for moral responsibility. Since I have already rejected the adding of libertarian objective conditions (4) and (5) above to strengthen the type of causal responsibility required to make it sufficient for moral responsibility, I shall maintain that causal responsibility is only necessary but not sufficient for moral responsibility for two reasons, one epistemic and the other normative. The epistemic reason is the difficulty in knowing all the relevant causes operating on persons, actions and their consequences. Even if we are able to establish whether persons exemplifying the relevant intentional states with the relevant reasons as content cause ('make a difference to', 'counterfactually support', 'contribute toward', or 'are INUS conditions of') actions, and that actions in turn cause ('make a difference to', 'counterfactually support', 'contribute toward', or 'are INUS conditions of') consequences, the difficulty of establishing to what degree persons exemplifying the relevant intentional states with the relevant reasons as content are morally responsible for their actions that result in consequences remains. This brings us to the normative reason. Causal responsibility provides only the objective (external) conditions for moral responsibility attribution. But moral responsibility attribution involves more than that, as it is a normative practice that also involves our reactive attitudes based on our shared values and norms. These are the subjective (internal) conditions for moral responsibility attribution. The normative limitation is that it is possible to have differences in shared values and norms, or differences in the subjective (internal) conditions for moral responsibility

attribution across communities and cultures, or even within a community or culture. Perhaps both the objective (external) conditions and subjective (internal) conditions are jointly sufficient for moral responsibility attribution but the limitations in establishing them remain.

The aim of this essay is not to give an account of the subjective (internal) conditions for moral responsibility ascription, but to argue for the necessity of the objective (external) conditions. The main reason for the significant relevance of the objective (external) conditions lies in the notion of action itself. Action necessarily involves bodily movement and intentional action necessarily involves voluntary bodily movement (or restraints). It is common knowledge that bodily movement, or even voluntary bodily movement (or restraints), have physical causes. To ascribe moral responsibility to persons exemplifying the relevant intentional states with the relevant reasons as content for their actions, they must figure at least as a partial cause of ('make a difference to', 'counterfactually support', 'contribute toward', or 'are INUS conditions of') their bodily movements, especially their voluntary bodily movements (or restraints). If not, the person, his intentional states, and their content become treated as mere epiphenomena that 'do not make differences to', 'do not counterfactually support', 'do not contribute towards', or 'are not INUS condition of' his actions, thereby making the normative practices of moral responsibility ascription mere social constructs formed by our shared customs and habits.

**References:**

Campbell, Joseph Keim (2011) *Free Will* (Polity Press)

Dowe, Phil (2004) ‘Causes are Physically Connected to their Effects: Why Preventers and Omissions are not Causes’ in *Contemporary Debates in the Philosophy of Science* (Blackwell Publishing)

Eshlemen, Andrew (2009) ‘Moral Responsibility’, *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/entries/moral-responsibility

Frey, R. G. (1978) ‘Causal Responsibility and Contributory Causation’, *Philosophy and Phenomenological Research Vol. 39 No. 1 pp. 106-119*

Hart, H.L.A., Honore, Tony (1985) *Causation in the Law 2nd Edition* (Oxford University Press)

Lewis, D. (1993): 'Causation' in *Causation* ed. Ernest Sosa & Michael Tooley (Oxford University Press)

Mackie, John L. (1993) ‘Causes and Conditions’ in *Causation* ed. Ernest Sosa & Michael Tooley (Oxford University Press)

Mckenna, Michael (2008) ‘A Hardline Reply to Pereboom’s Four-Case Manipulation Argument’, *Philosophy and Phenomenological Research Vol.76 pp.142-59*

Moore, Michael (2009) *Causation and Responsibility* (Oxford University Press)

Nagel, Thomas (2009) ‘Freedom’ in *Free Will 2nd Edition* ed. Gary Watson (Oxford University Press)

Pereboom, Derk (2009) ‘Determinism *al Dente*’, *Free Will* ed. Derk Pereboom (Hackett)

Schaffer, Jonathan (2004) 'Causes Need Not Be Physically Connected to their Effects: The Case for Negative Causation' in *Contemporary Debates in the Philosophy of Science* (Blackwell Publishing)

Smart, J.J.C. (2009) 'Free Will, Praise, and Blame' in Free Will 2nd Edition ed. Gary Watson (Oxford University Press)

Strasser, Mark P. (1992) *Agency, Free Will, and Moral Responsibility* (Hollowbrook Publishing)

Strawson, Peter (2009) 'Freedom and Resentment' in *Free Will 2nd Edition* ed. Gary Watson (Oxford University Press)

Thomson, J.J. (1987) 'The Decline of Cause', *The Georgetown Law Journal Vol.76 pp.17-150*

Trammell, Richard (1975) 'Saving Life and Taking Life', *The Journal of Philosophy Vol.72 No.5 pp.131-137*

Williams, Garrath (2009) 'Responsibility', Internet Encyclopedia of Philosophy (IEP) http://www.iep.utm.edu/responsi/

Wilson, H. Van Rensselar (1958) 'On Causation' in *Determinism and Freedom* ed. Sidney Hook (New York University Press)

# CHAPTER 6: REASONS, AGENCY, AND RESPONSIBILITY – A DEFENCE OF SOFT COMPATIBILISM

## Introduction

The thesis of soft compatibilism holds that a constrained notion of freewill is compatible with the thesis of both causal determinism and causal indeterminism, and that a constrained notion of freewill is compatible with a less robust sense of retrospective moral responsibility. This thesis of compatibilism is described as soft because both the notions of will and moral responsibility are 'softened' to make them compatible with one another as well as with either causal determinism or causal indeterminism. It denies that free will can be established by either causal determinism or causal indeterminism, but affirms that moral responsibility can be established by either causal determinism or causal indeterminism, as long as moral responsibility does not require the ultimacy condition. The thesis does not assume the truth or falsity of causal determinism or causal indeterminism, but holds that causality (whether deterministic or indeterministic) is necessary for moral responsibility. It does not identify causality as the main problem of the free will debate. Rather, it does identify the ultimacy condition (as attached to causal responsibility and moral responsibility) as the main problem. This thesis holds that once the ultimacy condition is rejected as necessary for causal responsibility and moral responsibility, causality can be shown to be compatible with moral responsibility if not free will. It is based on the following premises: (1) reasons (or teleological) explanation is a form of causal explanation, (2) agent (or substance) causation is a form of event causation, (3) moral responsibility does not require the ultimacy condition, and (4) causal responsibility grounds moral responsibility. A summary of the salient arguments in support of these premises will be given below.

**Reasons (or Teleological) Explanation is a form of Causal Explanation**

The first premise is that the agents' reasons are explanations of action and that reasons explanation is a form of causal explanation. And I attempted to defend a version of the causalist account (where reasons do figure as causes) against the non-causalist account (where reasons do not figure as causes) of reasons explanation. The salient argument against the non-causal account is that most morally significant action involves bodily movement or restraint, and voluntary action involves voluntarily bodily movement or restraint. The non-causal account does not explain the link between reasons and voluntary bodily movement or restraint. On the causalist account, reasons do make a difference, counterfactually support, contribute towards, or are INUS conditions of voluntary bodily movement or restraint. But on the non-causalist account, reasons do not make a difference, do not counterfactually support, do not contributes towards, or are not INUS conditions of voluntary bodily movement or restraint. Adequate explanations of voluntary bodily movements or restraints require ontological (causal) conditions and not just epistemological (truth) conditions. The causalist account is able to provide both conditions whereas the non-causalist account is able to provide the latter but not the former. Moreover, causal explanations and teleological explanations of action need not be mutually exclusive. The agent's ends, goals, purposes, or reasons (as used in teleological explanations) of his action can figure as part of the antecedent causal conditions (as used in causal explanations) of his action. They do so by being the content of the intentional states possessed by the agent when he engages in the process of practical reasoning. While the causalist account can accommodate both the causal and teleological aspects, the non-causalist account can only accommodate the teleological aspect.

**Agent (or Substance) Causation is a form of Event Causation**

The second premise is that events involving agents are causes of action and that agent causation is a form of event causation. And I attempted to defend a version of the event causalist account (where agents acting for reasons can be conceived as events) against the agent causalist account (where agents acting for reasons cannot be conceived as events). The salient argument against the agent causal account is its unintelligibility. The first version of the agent causal account combines the sole uncaused agent-cause and non-causal reasons as the explanation of action. But since the sole agent-cause is uncaused and reasons make no difference, do not counterfactually support, do not contribute towards, or are not INUS conditions of action, nothing explains why an agent does what he did. The second version combines the sole uncaused agent-cause with reasons as elevators of objective propensities and tendency-conferring states. But if reasons are states that elevate objective propensities and confer tendencies, then they are antecedent causal conditions (that make a difference, counterfactually support, contribute towards, or are INUS conditions) after all and the uncaused agent-cause cannot be the sole cause. And if the uncaused agent-cause is the sole cause, then reasons cannot be states that elevate objective propensities and confer tendencies (that are supposed to make a difference, counterfactually support, contribute towards, or are INUS conditions). The third version of the agent causal account combines an uncaused agent-cause with event-causes as co-causes. But even if the event co-causes explain why an agent does what he did, the uncaused agent co-cause remains unintelligible. So, the third version is unstable and it faces the dilemma of either rejecting the agent co-cause as redundant (hence collapsing into an event causal account), or rejecting the event co-causes as genuine causes (hence collapsing into the first version). All three versions of the agent-causal account are based on the more general substance causation, whereby only the

substance (object) figures as a cause, but not its states (or properties). By excluding states (or properties) as causes, substance causation cannot explain why, when, and how state-less (or property-less) bare substances can cause events to happen. Applied to action, this means that the agent (as substance) is the most significant if not the sole cause of action, and the intentional states and reasons as their content play very little or no causal roles. And the agent causal account cannot explain why, when, and how state-less (or property-less) bare agents can cause actions to happen. In contrast, my version of the event causal account holds that the substance (or object) and the states (or properties) it exemplifies (or instantiates) at certain times (or durations or time) as a whole figures as an event-cause. It can easily explain why, when, and how event-causes cause event-effects to happen. Applied to action, this means that the agent (as substance), the intentional states he exemplifies or instantiates, as well as the reasons as the content of his intentional states, as a whole, is the cause of action. And my event causal account can easily explain why, when, and how event-causes cause actions to happen. Hence, it is more intelligible to conceive of causes of action as event-causes involving agents or agent involving event-causes, rather than to conceive them as bare substances.

**Moral Responsibility Does Not Require the Ultimacy Condition**

The third premise is that retrospective moral responsibility does not require the ultimacy condition and that the ultimacy condition is both naturalistically implausible and normatively unnecessary. And I attempted to defend retrospective moral responsibility without the ultimacy condition from the hard incompatibilists as well as attack retrospective moral responsibility with the ultimacy condition of the libertarians. The reason ultimate causal responsibility is naturalistically implausible is because it requires the notion of uncaused self-

causation (causa sui), that is both logically and empirically implausible. Logically speaking, to cause the way our selves are (of which our character, personality, motivational structure follow) already presupposes that our selves are already existing in some way (having some character, personality, and motivational structure). Empirically speaking, we do not have complete or full control over the circumstances that affect the development of our character, personality, and motivational structure. A simple argument can now be given against the ultimacy condition based on the impossibility of ultimate causal responsibility: ultimate moral responsibility entails ultimate causal responsibility, and since ultimate causal responsibility is impossible, ultimate moral responsibility is impossible. Here, some incompatibilists may want to give up the stronger thesis in favour of the weaker thesis 'causal responsibility grounds moral responsibility'. But if they do, then their account has no distinct advantage over a compatibilist account, which can easily accommodate the weaker thesis. The stronger notion of moral responsibility is normatively unnecessary because there is no justification for attributing moral responsibility to a person for circumstances that affect the development of his character, personality, and motivational structure that are beyond his control. Attributing a weaker notion of moral responsibility by taking into account his capacity for practical reasoning, process of practical reasoning, and his relevant intentional states and reasons that cause his actions seems adequate for our normative practices. If these normative practices involve the ultimacy condition, then very little if any of excusing conditions and mitigating factors are accepted. This is because we would be very reluctant to accept excusing conditions or mitigating factors if we accept that we are ultimately responsible (in both the causal and moral sense) not only for all our choices and actions, but for our wills, character traits, and motives as well. The ultimacy condition ends up making us morally responsible for more than we actually deserve. Another simple argument can now be given against the ultimacy condition based on the non-necessity of ultimate moral

responsibility: ultimate moral responsibility entails ultimate causal responsibility, and since ultimacy is not required for moral responsibility, ultimacy is not required for causal responsibility. However, the rejection of the ultimacy condition is not a rejection of causal responsibility and moral responsibility. For these arguments do not rule out weaker theses like the following: moral responsibility entails causal responsibility, if we are morally responsible, then we are causally responsible, and if we are not causally responsible, then we are not morally responsible.

**Moral Responsibility is Grounded on Causal Responsibility**

The fourth premise is that causal responsibility grounds retrospective moral responsibility, where moral responsibility can be justifiably attributed to the agents' actions and their consequences only when the agents' exemplification of the relevant intentional states with the relevant reasons as content are causally responsible for them. And I attempted to defend the premise against various objections. The salient objection against this premise is that our theoretical convictions like those about causal theses are irrelevant to our practices of moral responsibility ascription. These practices can only have subjective or internal justifications based on our shared customs and habits, not objective or external justifications based on our theoretical convictions. This is because our shared customs and habits are too deeply entrenched in us to be changed by our theoretical convictions. The reply to this objection is that since moral responsibility is ascribed to the agent who performed an action, morally significant action involves voluntary bodily movement or restraint, and voluntary bodily movements or restraints have causes, then our theoretical convictions like causal theses are relevant to our practices of moral responsibility ascription. While it may be true that our shared customs and habits are too deeply entrenched in us to be changed by our theoretical

convictions, it is not true that our theoretical convictions play no role in our practices of moral responsibility ascription. To ascribe moral responsibility to agents exemplifying the relevant intentional states with the relevant reasons as content for their actions, these reasons must figure at least as a partial cause of ('make a difference to', 'counterfactually support', 'contribute toward', or 'are INUS conditions of') their voluntary bodily movements or restraints. If not, agents, their intentional states and their content become treated as mere epiphenomena that 'do not make differences to', 'do not counterfactually support', 'do not contribute towards', or 'are not INUS condition of' their actions, thereby making the practices of moral responsibility ascription mere social constructs formed by our shared customs and habits. Moral responsibility ascription may not need 'obscure or panicky metaphysics', but this does nothing to show that moral responsibility ascription does not need metaphysics at all.

**Comparing Soft Compatibilism to Other Positions in the Free Will Debate**

Causal determinism (or causal indeterminism) may or may not be true, and it remains an unresolved empirical question. Claiming the truth or falsity of either causal determinism or causal indeterminism constitutes a theoretical commitment to an unresolved empirical question. Both libertarian (absolute) free will and libertarian (ultimate) moral responsibility requires the falsity of causal determinism (or truth of causal indeterminism), and hence libertarians are theoretically committed to these unresolved empirical claims. Both the hard and soft determinism require the converse of the libertarians' claims: the truth of causal determinism (or falsity of causal indeterminism), and hence both hard and soft determinists are too theoretically committed to these unresolved empirical claims. These commitments constitute a theoretical disadvantage for the libertarians, hard determinists, and soft

determinists. In place of ultimate moral responsibility, soft compatibilism instead accepts an account of moral responsibility that does not require the truth or falsity of either causal determinism or causal indeterminism, but that is consistent with either of them. Hence, soft compatibilism is not theoretically committed to these unresolved empirical claims, and the absence of such commitments constitutes a theoretical advantage for the soft compatibilists. Soft compatibilism does not even require the truth of universal causality – the thesis that every event has a cause or that all events are caused, but it does require that every morally relevant event has a cause (whether deterministic or indeterministic). It is incompatibilist about libertarian free will and causal determinism as well as about libertarian (ultimate) moral responsibility and causal determinism, but it is compatibilist about moral responsibility (without the ultimacy condition) and causality (whether deterministic or indeterministic). And it denies libertarian free will and libertarian (ultimate) moral responsibility whether or not causal determinism is true.

However, hard incompatibilism and other forms of compatibilism that are not committed to the truth or falsity of causal determinism or causal indeterminism share that same theoretical advantage as soft compatibilism. So, what theoretical advantage does soft compatibilism has over hard incompatibilism? The salient one is that while the soft compatibilist preserves retrospective moral responsibility (without the ultimacy condition), the hard incompatibilist eliminates it. And what theoretical advantage does soft compatibilism has over other forms of compatibilism that are not committed to the truth or falsity of causal determinism or causal indeterminism? Compared to the form of compatibilism inspired by Peter Strawson's work, soft compatibilism takes the relevance of causality seriously, and it thereby provides an objective or external ground for moral responsibility ascription. And compared to the form of compatibilism inspired by Harry Frankfurt's work, soft compatibilism identifies and

eliminates the ultimacy condition as the problem, and it thereby makes the notion of retrospective moral responsibility more plausible and defensible for the compatibilists. These considerations seem to favour soft compatibilism over all the other positions discussed here.